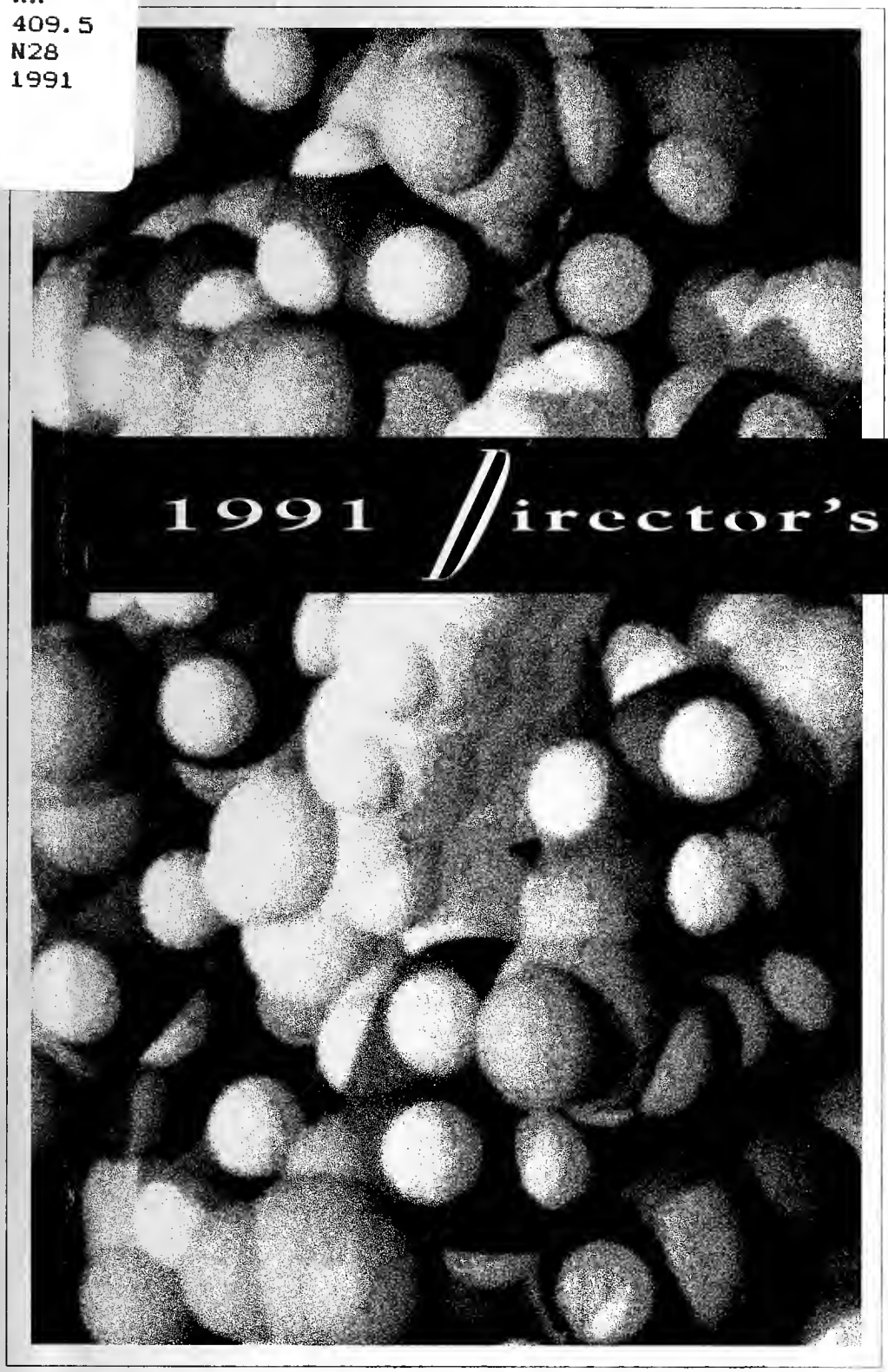
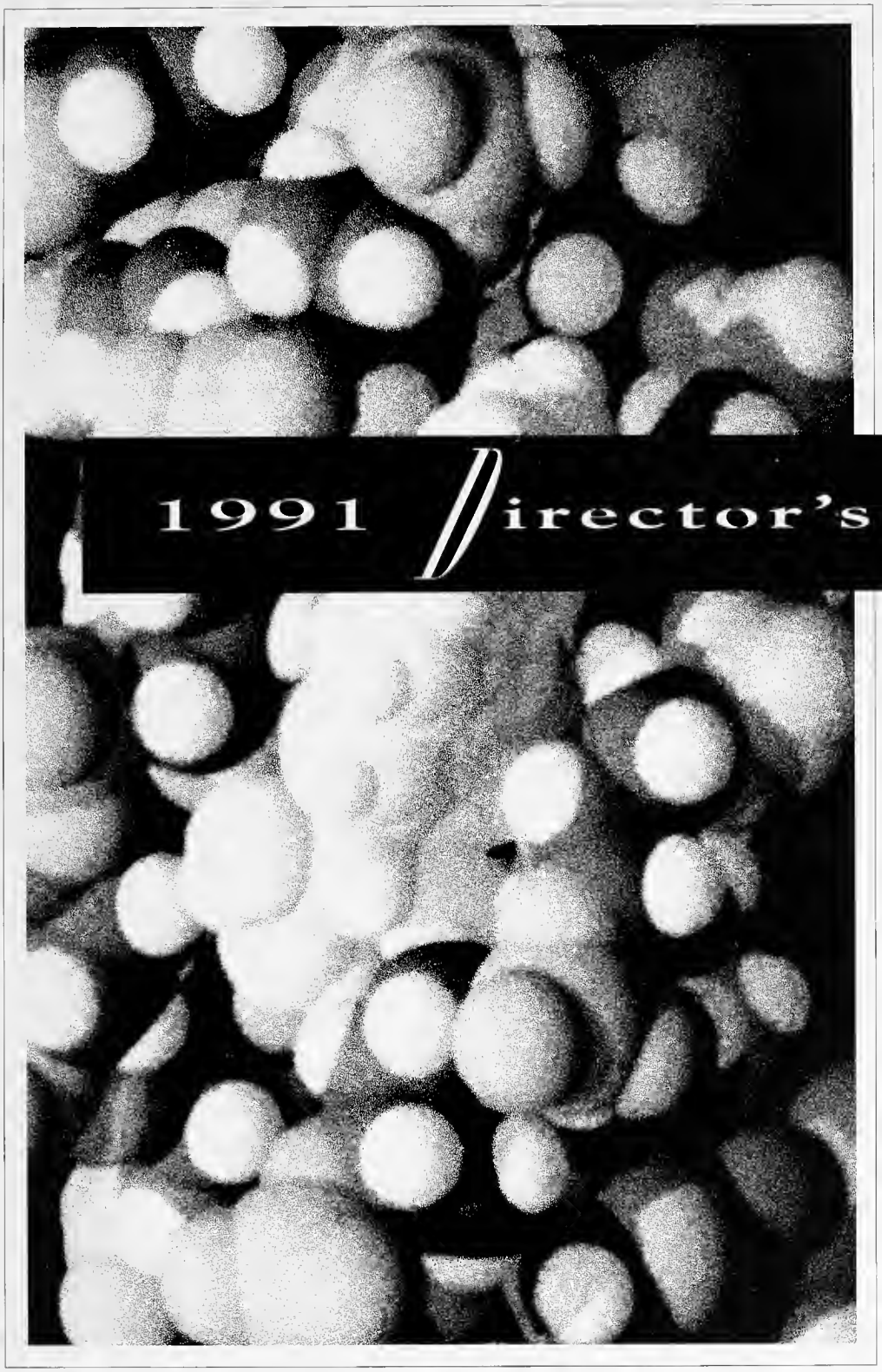


1991 Director's Report

Division of Computer Research and Technology

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
The National Institutes of Health

Cover (full view):
Computer-generated image of a carcinogen (benzopyrene-diolepoxide, or BPDE) bound in a DNA double helix. (DCRT)



1991 Director's Report

Division of Computer Research and Technology

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
The National Institutes of Health

The Division of Computer Research and Technology (DCRT) has primary responsibility for incorporating the power of modern computing and networking into the biomedical research programs and administrative procedures of NIH. DCRT serves as a scientific and technological resource for the Public Health Service and for other Federal organizations.

DCRT programs focus on three primary activities: conducting research, developing computer systems, and providing computer facilities.

The ***Physical Sciences Laboratory*** *conducts research in biomathematics and both theoretical and experimental biophysics to explain biological phenomena in terms of chemistry and physics at the subcellular molecular levels.*

The ***Computer Systems Laboratory*** *consults and collaborates with NIH investigators in the development of computer-based systems for laboratory and clinical applications, and conducts computer science and engineering research and development directed toward the application of new computer technologies to NIH programs. Special areas include advanced laboratory workstations, central support of distributed computing, image processing (clinical and laboratory), massively parallel scalable high-performance supercomputing, support for the Human Genome Initiative Expressed Sequence Tags, and support for analytical flow cytometry.*

The ***Laboratory of Statistical and Mathematical Methodology*** *provides consultation and collaboration in computational, biostatistical, and biomathematical aspects of data analysis; supports software systems and packages to perform these analyses for the biomedical community; and conducts independent research in biostatistics and biomathematics with relevance to laboratory research, clinical trials, and other biomedical applications.*

The ***Laboratory of Applied Studies*** *applies biomathematics and computer sciences to biomedical problems such as ECG analysis, modeling for the evaluation of physiological systems in health and disease, and data analysis problems in laboratory experimentation.*

The ***Network Task Group*** *develops and applies state-of-the-art computer communication technologies to support and enhance NIH's biomedical and clinical research programs. NTG implements the RESnet component of NIHnet (the NIH on- and off-campus internetworking facility).*

The ***Personal Computing Branch*** *provides guidance, support and training to scientists and administrators throughout NIH in the effective use of personal workstations, local area networks and associated office laboratory automation technology. PCB works closely with other DCRT labs and branches to monitor these rapidly changing technologies with a view toward the technical requirements of the user community so that these needs are accommodated in the development and support provided by DCRT.*

The ***Data Management Branch*** *serves as the central systems analysis, design and pro-*

gramming resource for data processing projects relating to scientific, technical, management, and administrative data. DMB oversees the NIH Administrative Data Base which integrates the administrative and financial data of the intramural program. Its recently formed New Technology Analysis Section will analyze and select new database management approaches and develop methodologies to ease their use across multiple platforms.

The **Computer Center Branch** *operates and maintains the NIH Computer Utility and the NUnet component of NIHnet. CCB designs and develops software and networking facilities to provide flexible services for scientists and administrators at NIH. It also provides extensive personal support and documentation for more than 17,000 computer and network users, and its Computer Training Program offers many popular courses and seminars. Additionally, CCB provides high-performance scientific computational services via a Convex superminicomputer accessible through NIHnet.*

The **Office of the Director** *provides overall program direction for DCRT, and serves as a central NIH focus for development of policy for automated data processing. In addition, the Office sponsors research and development work in computational biosciences (e.g., computational structural biology, computational molecular biology and genomics).*

From the Director

This year was a momentous one for DCRT. Gradual evolution had been occurring in the Division for 25 years, but 1991 marked the beginning of a renaissance. For the first time since the organization's beginning in 1966, we addressed the very essence of our existence, taking methodical steps to ask such basic questions as:

- What is our mission?
- What are our goals?
- What are our priorities?
- How can we make the best use of our finite resources?
- What are the major challenges facing the Division?
- How can we provide the appropriate range and quality of services in the most cost-effective manner?

Perhaps most importantly, "What can we do to provide the very best research and development, resources and service in support of the now clearly enunciated mission, goals, and priorities of the NIH strategic plan?"

Two of the most important days of the year were August 1 and September 3, when we held two DCRT-wide retreats. Personnel representing all branches, laboratories, specialties and aspects of DCRT met to discuss vital questions and to examine cross-cutting collaboration and communication among research, development, service, support, and administrative staff. This kind of communication had rarely, if ever, taken place. These events closely preceded the first meeting of a DCRT *ad hoc* advisory council, consisting of 10 expert consultants representing almost every aspect of computing and communication in biomedical research. In turn, the Division participated in the September 10-11 pan-NIH strategic planning conference. DCRT staff, in concert with a strategic planning consultant, coordinated the Division's contribution, preparing position papers on structural biology, biotechnology, molecular medicine, research infrastructure, and communications and information flow. Within the next few months, after some further refinement, our DCRT strategic plan will be sufficiently developed to serve as the foundation for our endeavors for many years to come. We are ready to take concrete steps toward implementation.

The effort has already begun. In FY91, we initiated the process of systematic peer review of all Division activities, engaging the services of nearly 30 notable specialists in biomedical and computer science. Four panels of these experts examined networking, structural and molecular biology, and our activities in the fields of biomathematics, biostatistics, biophysics, and image processing (imaging sciences). These reviews provided an objective, independent database upon which we can rest our strategic plan as well as immediate operational decisions.

High on the list of our priorities was careful attention to planning for the revision of Information Resources Management (IRM) both at DCRT and NIH. Many hours were profitably invested by our staff in crafting new approaches to this complex issue, which is becoming an increasingly important aspect of life at NIH and throughout all Federal agencies. We are confident that our efforts in this area will help to give NIH an IRM program befitting its status as the Nation's premier biomedical research institution.

Another major area of activity in FY91 was the development of a plan for our active participation in the High Performance Computing and Communications (HPCC) initiative of the Federal Coordinating Committee on Science, Education, and Technology. Our detailed agenda for intramural research will allow us to apply the massively parallel, scalable supercom-

puter to a wide variety of vitally important, computationally intensive problems in biomedical research, including image processing, genetic database searching, gene linkage analysis, structural biology, and approaches to the grand challenge of the protein-folding problem. Already, the power of parallel computing has been used to determine the three-dimensional structure of clinically important pathogenic viruses by computerized reconstruction from electron micrographs. In addition, programs to facilitate molecular dynamics and calculation of solvent-accessible surface area are being implemented.

Our parallel computing capacity has grown from 16 to 64 to 128 i860 processors, each one nearly as powerful—in terms of megaflops—as a mainframe computer. Within a few years, we expect to have computers with thousands of processors in parallel, making it possible to attack problems that could only have been dreamed of years ago. At present, DCRT is the only NIH component with active ongoing research using this rapidly evolving technology. We shall provide leadership in this revolutionary new field, making it available to all of the other Institutes, Centers, and Divisions. Working together with the National Library of Medicine, the National Center for Research Resources and the National Cancer Institute's Advanced Scientific Computing Laboratory, we have developed a detailed plan for high-performance computing. This activity is likely to provide a new source of funding, and facilitates our collaboration, cooperation, and communication with other government agencies, including the Defense Advanced Research Projects Agency, the Department of Energy, the National Science Foundation, the National Aeronautics and Space Administration, and the National Institute of Standards and Technology.

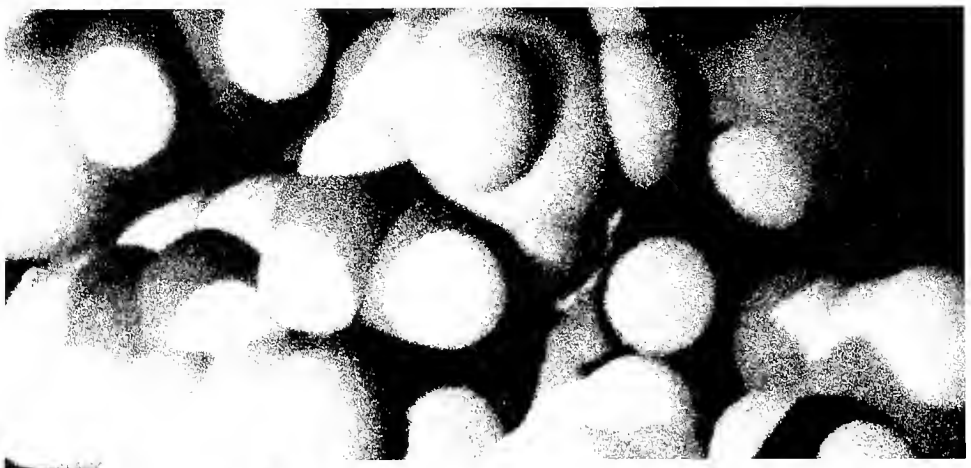
DCRT has taken a leadership role in the "modernization" (viz., complete redesign and reimplementation) of the CRISP and IMPAC databases that are vital to the operations of the entire NIH extramural community. We are working closely with the Division of Research Grants and a committee of policy and technical representatives from all of the Institutes, Centers, and Divisions to provide a test of three potential solutions in this multiyear, multimillion-dollar strategic project. The chosen solution must be flexible enough to serve the changing needs of the research community for 10-20 years to come, and to adapt as hardware, software, networks and user requirements evolve. This massive modernization project already involves the Division's Computer Center Branch, Data Management Branch, and Personal Computing Branch; other DCRT components are likely to become progressively more involved as truly interoperative client-server facilities become available. Technology is evolving so rapidly that "technology tracking" has become a major Division activity. Databases are also being applied to facilitate research in molecular biology and genomics. DCRT and the Clinical Oncology Program of the National Cancer Institute have begun a collaboration to develop database systems to facilitate clinical research in such areas as data collection, data analysis reporting, clinical monitoring, and the preparation of reports for the FDA.

In the next year we shall complete our peer review of all DCRT activities, finalize the initial DCRT Strategic Plan, and proceed aggressively with its implementation.We can expect many changes in DCRT as we provide leadership in the ever-changing fields of computing and biomedical research. Based on our peer review conducted in June 1991, we are already planning the creation of a Computer

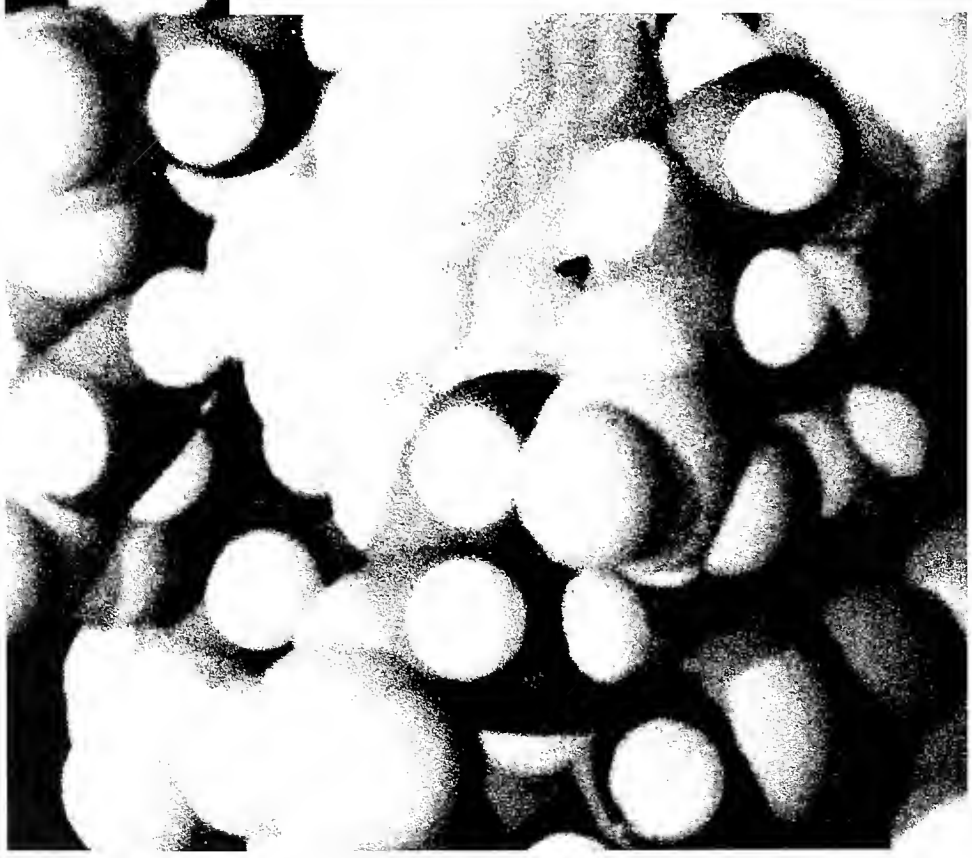
Networking Branch to enhance our ability to move forward rapidly in this vital area. Based on our fourth peer review (August 1991), we are moving forward with plans for a medical image processing group. We have already strengthened our collaborative relationships with the Clinical Center, the National Center for Research Resources, and nearly all of the Institutes by providing leadership and expertise necessary for the MRIPS (Multimodality Radiology Image Processing System) project. We are committed to the inauguration of a Scientific Computing Resource Center within the next year to provide enhanced consultation services to the NIH scientific community.

We have dramatically improved our capacity planning program, using computer modeling techniques to assist in the design and configuration of our mainframe facilities. We shall do everything possible to ensure that all operations are as cost effective as possible.

Contents



CB



Computational Biosciences

Computational Molecular Biology Section

Peter C. FitzGerald, Ph.D.
George S. Michaels, Ph.D.

The Computational Molecular Biology Section (CMBS) is comprised of scientists with a background in both molecular biology and computational science. The section is chiefly responsible for providing support for NIH intramural scientists interested in applying computer-based tools to the collection and analysis of DNA and protein sequence data. CMBS offers consultation, support, and guidance in computational biology to any interested NIH scientist. In addition to its primary role as a support entity, the CMBS pursues research in computational molecular biology.

Research within the CMBS has been directed toward the development of new methods for manipulating DNA and protein sequence data to facilitate the analysis of genomic organization. Several projects have been initiated that are designed to take advantage of distributed databases and computational facilities. The projects include the establishment of a network-accessible zinc finger gene database, molecular modeling and statistical analysis of zinc finger domain/nucleic acid interactions, the formation of an internet collaborative working group to design new prototype informatics systems for the logical manipulation of chromosomal map information, and the development of new tools for the analysis of DNA and protein sequences.

FY91 Accomplishments

The CMBS currently supports the Genetics Computer Group (GCG) sequence analysis software, as implemented under Unix on the DCRT Convex C240. More than 200 NIH research staff use the package; some 80 scientists were trained in the use of GCG during the first half of 1991. The CMBS is responsible for maintaining copies of the major DNA and protein sequence databases on the NIH Convex System where they can be accessed by NIH scientists in a number of different formats.

In an effort to promote and facilitate DNA and protein sequence analysis on personal computers, the CMBS negotiated an NIH-wide site license with the producers of the "MacVector" program, and distributed 50 copies of this software to NIH laboratories.

During the past year the section has been actively involved in expanding the variety and quality of programs and facilities available to scientists in the area of computational molecular biology. These objectives have been pursued for a wide variety of computer architectures, including the NIH Convex C240, Macintosh and IBM personal computers and a variety of Unix workstations. In addition to direct user assistance for supported applications, the CMBS has provided scientific consultations to individual NIH scientists interested in initiating computer analysis of sequence data; interpreting the biological significance of computer-based analyses of data; and designing future biological experiments based on computer analysis of existing data.

In its support role, the CMBS interacts with NIH scientists from all ICDs, including individuals from the main NIH campus in Bethesda and from NIH satellite facilities. In addition, due to its role of supporting DNA and protein sequence analysis on many different computer types, the CMBS has close ties with many other groups within the DCRT; most significantly with the Convex System staff of the Computer Center Branch, the Personal Computing Branch, the Computer Systems Laboratory, and the DCRT/CCB Training Unit.

In collaboration with researchers at the Medical Research Council, Cambridge, UK, a collection of zinc finger peptide sequences containing both published and unpublished data has been assembled. New

finger domain sequences can be electronically mailed to DCRT where they are added to the database; a "FASTA" search is performed, and a report is returned to the investigator. Alignments are not supplied in the report, but a histogram of the statistical distribution of the search results and a listing of the potential scores are included. If a match to an unpublished sequence occurs, then the name and contact information of the submitting author is returned so the concerned parties may correspond with each other. To date the collection contains 147 gene entries accounting for 1,257 finger domains.

The development of integrated tools for the analysis and manipulation of genomic information has been the goal of an Internet-based "collaboratory" of computer scientists and computational biologists who have been working together to apply logic programming tools to the analysis of genomic information. The specific aims for this work are: (a) to establish the minimal logical criteria necessary to describe genomic map data, (b) to apply such criteria to model chromosomes, and (c) to develop a framework of computational analysis facilities that can manipulate aligned genomic sequencing and mapping data. A series of rapidly prototyped deductive relational database query systems have been developed in Prolog that allow analysis of DNA sequence information while taking advantage of the interrelated physical, hybridization, and genetic mapping data.

Also under development is a Multiple Map Analysis Visualizer (MMAV) tool that allows interactive evaluation of aligned DNA sequence pattern analysis data. The tool runs on any SunSPARCstation using an OSF X11-based window manager. Input to the program is in the form of simple text files, allowing the incorporation of positional information from most sequence analysis systems.

Future Plans

In the coming years, DCRT plans to expand both research and support services in the areas of molecular biology and genomics. Such expansion will encompass a crucial role in the design, development, and support of a new Scientific Computing Resource Center and an increase in support provided to NIH scientists. The CMBS plans to increase the number and variety of sequence analysis programs available on the NIH Convex C240 System. The CMBS will provide additional support for the use of these applications and expand its training efforts. In 1992 the CMBS will be involved in the implementation of one or more options for DNA and protein sequence analysis on Unix workstations. The CMBS will also continue its efforts to improve access to sequence analysis software through continued "site licensing" of commercial software packages which execute on personal computers and workstations, and through gathering and distribution of public domain code.

Future development of distributed network database tools of DNA and protein sequence analysis will concentrate on the development of specialized prototype databases to be used for individual analysis methods. For example, DNA sequence-repeat databases will be used to classify the distribution of these repeats in long human genomic sequences. The collaboratory workshops will pursue interoperability standards for queries and exchange of genomic information.

Homology modeling of zinc finger domains will be used to identify potential DNA binding sequences for individual domain structures. We expect this work to extend the understanding of how the superfamily of potential transcriptional regulatory genes that code for proteins containing the zinc finger motif are involved in embryonic pattern formation.

Research Projects

Logic Programming-Based Query System for Chromosomal Information

G. Michaels, Ph.D., G. Dunham, R. Taylor with R. Overbeek, Ph.D., (Argonne National Laboratory, Chicago, Illinois); D. Hartl, Ph.D., T. Kazic, Ph.D., (Washington University, St. Louis, Missouri); K. Rudd, Ph.D., (NCBI/NLM); C. Smith, Ph.D., (University of California, Berkeley); K. Yoshida, Ph.D., (Lawrence Berkeley Laboratory, Berkeley, California); D. Zawada, (Argonne National Laboratory, Chicago, Illinois)

We have organized a computational biology "collaboratory" of scientists, both computer scientists and biologists, that gathers periodically to work on the development of a new set of integrated tools for the manipulation of genomic information. Development of these informatics tools and assembly and analysis of the integrated data continues over the Internet between meetings. The initial objective of this research was to establish the minimal criteria necessary to describe genomic map data that may be logically manipulated. The result of this collaborative effort has been the development of three prototype deductive database systems:

- the *E. coli* chromosome query system that contains information provided by K. Rudd at NCBI/NLM for aligned DNA sequences, a high-resolution physical map, identified structural genes, and an aligned phage map
- at Lawrence Berkeley Laboratory, an object-oriented deductive database for the collective information of over 30 different types of genetic and physical map data for human chromosome 21
- a relational representation of the information provided by D. Hartl for the aligned *Drosophila* chromosome band and Yeast Artificial Chromosome (YAC) hybridization data.

The aligned chromosome information for each of these prototypes may be viewed using a common graphical display program developed at the Argonne National Laboratory for the collaboratory.

A common feature of each of these prototype data representation systems has been use of the logic programming language Prolog. It is through Prolog that we can rapidly develop complex queries of the integrated data that take advantage of the complicated interrelationships inferred.

Additionally, we developed here in the CMBS a Multiple Map Analysis Visualizer (MMAV) tool. This program allows the interactive evaluation of the several aligned DNA sequence pattern analysis data. The tool runs on any SunSPARCstation using a OSF conforming X11-based window manager. Input to the program is in the form of simple text files, allowing the incorporation of positional information from most sequence analysis systems. The visual display of aligned maps of user-selectable multiple analysis data facilitates thoughtful consideration of the spacial relationships between the occurrences of sequence features.

Prototype Genome Informatics Systems

G. Michaels, Ph.D., R. Feldmann with S. Uchida, Ph.D., (Institute for New Generation Computer Technology, Tokyo, Japan); R. Overbeek, Ph.D., (Argonne National Laboratory, Chicago, Illinois); K. Yoshida, Ph.D., (Lawrence Berkeley Laboratory, Berkeley, California)

The analysis of genomic organization requires a facility that allows a biologist to logically manipulate the complex interrelated information thoughtfully. A collaborative working group has been established to explore new hardware and software technologies that may be applied to the analysis of genomic information. The goals for this group are to apply logic programming to develop prototype analysis and simula-

tion systems for biological systems. A series of collaborative visits and workshops has been used to define three research areas: 1) simulation of protein folding using rule-based methods, 2) logical manipulation of chromosomal map information, and 3) identification of sequence motifs.

Understanding Protein-Nucleic Acid Interactions

G. Michaels, Ph.D., S. Knisley, P. Munson, Ph.D., with I.B. Dawid, Ph.D. (NICHD/LMG); G. Jacobs (MRC, Cambridge, U.K.)

What is the structure and mechanism of interaction of the CC-HH finger domains with nucleic acids? This question has been addressed through a detailed structural analysis of the finger regions from several known nucleic acid binding proteins. A database of zinc finger proteins has been assembled for the purpose of statistical and structural modeling of the individual finger regions. The "Zinc Finger Database" was established to accumulate a complete collection of potential zinc finger gene sequences that could be rapidly searched by the research community to prevent duplication of efforts. This collection containing both published and unpublished gene sequence data is maintained on the DCRT Convex 240. New zinc finger gene sequences are electronically mailed to the NIH by investigators. On receipt, the data are added to the database and a FASTA search of the results is returned without the alignments. A histogram of the statistical distribution of the search results and listing of the potential scores is included. When a match to an unpublished sequence occurs, the name and contact information for the submitting author is returned so the concerned parties may correspond with each other. To date the collection contains 147 different entries.

Statistical analysis of these data has revealed five repeat classes of CCHH zinc finger domains ranging in length from 27 to 32 amino acids, and 22 different repeat patterns. Furthermore, compositional statistics of the largest class of domains, 29 amino acid repeat length, reveals a remarkable conservation of serine or threonine when there is an arginine or glutamine in the DNA binding region of the finger domain. A correlation between the nucleic acid sequence recognition has been established for some finger domains. Constrained molecular dynamics simulations suggest that there are two general folding motifs for the zinc finger domains. Domains with sequences consistent with the consensus structure can adopt folds similar to the published x-ray data for the zf268 protein/DNA co-crystal, while non-consensus structures adopt a flexible beta-loop conformation. Homology-based molecular models using the published x-ray crystallographic data are under construction. These homology-based structural models will be used for molecular docking experiments that explore the energetics of DNA sequence recognition.

Publications and Presentations

Graessman M, Michaels GS, Graessman A. Inhibition of SV40 gene expression using microinjected small antisense molecules, *Nucleic Acid Research* 1991;19:53-59.

Jacobs G, Michaels GS. Zinc finger gene database, *New Biologist* 1990; 2(6):583-584.

Kazic T, Michaels GS, Overbeek R, Zawada D, Dunham G, Smith CL. An integrated database of *E. coli* chromosomal information to support queries and rapid prototyping, AAAI Workshop on Applications for Genome Analysis, July 19, 1991.

Michaels GS, Kazic K, Overbeek R, Zawada D, Dunham G, Rudd K, Smith CL. Logic programming-based system for querying *E. coli* chromosomal information, Cold Spring Harbor, New York, May 8-12, 1991.

Molecular Graphics and Simulation Unit

Bernard R. Brooks, Ph.D.

The Molecular Graphics and Simulation (MGS) unit studies problems of biological significance using the theoretical techniques of molecular dynamics, molecular mechanics, modeling, *ab initio* analysis of small molecule structure, and molecular graphics.

The MGS unit conducts both collaborative and independent research projects with the Biophysics Laboratory, the Center for Biologics Evaluation and Research (CBER), FDA and with the NCI, NIMH, NIDDK, NHLBI and NIDR. Additional collaborations are active with the Boris Kidric Institute, Ljubljana, Yugoslavia; Eli Lilly, Indianapolis, Indiana; Star Technologies Inc., Sterling, Virginia; Georgetown University; and Harvard University. Group members also perform a service role at the NIH, supporting and encouraging the use of scientific computing as a research tool.

Molecular Dynamics Simulations of Biological Macromolecules

Research Involving AIDS Proteins

The MGS unit, in collaboration with the Biophysics Laboratory of the CBER, has been part of the NIH Intramural AIDS Targeted Antiviral Program since 1987. Important therapeutic targets under study include HIV-1 reverse transcriptase, HIV-1 protease, the HIV-1 envelope protein gp120, and the CD4 receptor protein found on certain host cells. Projects include:

- Modeling of Leucine Zippers Using Molecular Dynamics: Transcription Factor GCN4 and the Dimerization Domain of HIV-1 Reverse Transcriptase
- Molecular Dynamics Simulations of HIV-1 Protease Monomer in Solution.

Other Applied Research on Molecules of Biomedical Interest

Applied simulation research uses molecular dynamics simulations to predict function or structures of peptides and proteins, with application to specific biomedical goals, such as vaccine or antiviral agent development. Projects include:

- The Role of cAMP in the Stability and Activity of Catabolite Gene Regulatory Protein
- A Molecular Dynamics Simulation of the L-Alanine Crystal.

Basic Research

Basic research provides a better understanding of biochemical systems. Emphasis has been on simulations to analyze structure/function relationships and other properties of macromolecules. Projects have included computational drug design, design of transition state analogs, and *ab initio* calculations for determining reaction mechanisms. This type of research is also needed for the testing and evaluation of methods and models. Current projects include:

- Environmental Effects on Protein Dynamics
- Time Dependent Relaxation Behavior of Proteins: Interleukin-1b
- Harmonic Analysis of Large Systems
- Modeling and Simulation of the Lipid Bilayer
- Molecular Dynamics Simulation Studies of DNA: the B-Z Junction.

Development of Theoretical Methods for Studying Biological Macromolecules

Algorithms and Software

New theoretical techniques under development are often coupled with software and hardware development, involving the generation of new simulation techniques and the systematic testing and evaluation of methods. Projects include:

- Analysis of Hysteresis in Free Energy Perturbation Simulations
- Slow Growth Homology Modeling
- Development of Quantum Mechanical Potentials and Appropriate Simulation Algorithms for Use in Molecular Dynamics Simulations
- Excited State and Electron Transfer Processes in Biological Systems
- Semiempirical Hartree-Fock Calculations of Proteins.

Parameters and Models

Many of the parameter sets and models that are generally available are of the quality required for accurate simulation of macromolecular systems. Though parameter development could easily consume all of the MGS unit personnel and computing resources, we restrict our parameter development efforts to areas of primary interest where the existing parameter sets are inadequate. One such example is in the development of parameters for simple organic substituents to use in modeling lipids. Projects include:

- Development of van der Waals Parameters for Methylene and Methyl Groups
- Development of Parameters for Lipid Bilayer Simulations.

Development of Advanced Computer Hardware and Software

Software and hardware efforts include the development of microcode for commercial processors, and then design of specialized high-performance computers for specific computational chemistry needs.

Massively Parallel Computers

Development of methods and software to make productive use of parallel multiple instruction-stream, multiple data-stream machines for use in macromolecular simulations is underway. The initial approach has been to support a full feature version of CHARMM (CHemistry at HARvard Macromolecular Mechanics) using a straightforward approach to parallelization. Performance statistics indicate that this approach is probably optimal for the current hardware architecture.

The GEMMSTAR Project

The MGS unit has evaluated a project to design a high-performance, application-specific computer for molecular simulations in collaboration with the Computer Systems Laboratory (CSL) and Star Technologies Inc. with a Cooperative Research and Development Agreement (CRADA) under the auspices of the Federal Technology Transfer Act of 1987.

Due to the advent of high-speed workstations available at low cost, the effort to develop application-specific hardware has been reduced to the evaluation of computer architectures for macromolecular simulation.

Support Activities

The MGS unit is actively supporting molecular modeling and simulation needs at the NIH both through

consulting and formal training. Direct services provided by the MGS unit include:

- Research support and guidance for NIH scientists
- Provision of an NIH resource for short-term graphics and modeling needs
- Support for software packages on a variety of hardware platforms
- Examination and evaluation of new hardware
- Assessment of needs at the NIH and provision of policy recommendations to the DCRT management and other DCRT organizations
- Assistance to other DCRT sections in making their computational resources useful for the research needs of NIH.

Courses and Seminar Series

The following courses and seminars were offered by the MGS unit this year:

- CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations
- Usage and Applications of Molecular Quantum Mechanical Programs
- Molecular Dynamics for Problems in Structural Biology.

The MGS unit is also conducting an active seminar series for computational chemistry and runs a book review series, both of which are open to interested scientists.

Future Plans

In FY92, the MGS unit will continue to study the relationship between structure and function, and develop the theoretical analysis of inhibitor binding, and specifically study the mechanism of the HIV-1 protease. New methods are being developed for performing free energy perturbation calculations which allow bond connectivity to change during a simulation and have a higher accuracy than existing methods. Also under development are methods for treating solvent implicitly to provide for hydrophobic effects without the explicit inclusion of many water molecules; methods to properly treat electronic polarization in molecular dynamics simulations; and a more accurate flexible water model. Projects include:

- Modeling and Simulation of Winter Flounder Ice Inhibition Peptides
- Structural Analysis of T4 Lysozyme Mutants in the Harmonic Limit
- Cooperativity of Cystine Bridge Formation in Peptides and Proteins
- Molecular Dynamics Simulations on Staphylococcal Nuclease: Comparison with NMR Data
- Structural Characterization of a Heme:Myoglobin Adduct Using Molecular Mechanics
- New Methods for Long-Range Truncation of the Energy Potential
- Further Refinement and Examination of Free Energy Techniques
- Development and Use of a Polarizable and Flexible Water Model
- Molecular Dynamics Simulation Studies of DNA: Analysis of the Parameter Sets
- Conversion of Physical Models into Three-Dimensional Coordinates for Computer Analysis and Simulation
- Design and Analysis of Computational Methods Exploiting Massively Parallel Architectures.

The MGS unit will continue to be a resource for NIH and provide direct collaborative assistance, and will continue to give courses and to organize seminar series and book review series.

Research Projects

Molecular Dynamics Simulations of Biological Macromolecules

B.R. Brooks, Ph.D.; F.W. Carson, Ph.D.; I. Chandrasekhar, Ph.D.; D. Janezic, Ph.D.; R.J. Loncharich, Ph.D.; W.M. Southerland; P.J. Steinbach, Ph.D.
with R.M. Venable, R.W. Pastor, Ph.D. (FDA/CBER); M. Bansal, Ph.D.(Molecular Biophys. Unit, Indian Inst. of Science, Bangalore, India)

Applied theoretical research on AIDS proteins and other molecules of biomedical interest as well as basic research is in progress.

Molecular dynamics simulations of AIDS proteins involve projects directly related to the NIH Intramural AIDS Targeted Antiviral Program. The general goal is to understand binding interactions with HIV-1 proteins in order to facilitate the design of drugs which may interfere with the spread of the virus. Important therapeutic targets under study include HIV-1 reverse transcriptase, HIV-1 protease, the HIV-1 envelope protein gp120, and the CD4 receptor protein found on certain host cells. Projects include modeling of leucine zippers in GCN4 and HIV-1 reverse transcriptase, simulations of HIV-1 protease monomer in solution, analysis of CD4 fragments and derivatives, analysis of inhibitor binding to the active site of HIV-1 protease, and investigation of the mechanism of action of HIV-1 protease.

The analysis of the protease monomer simulation shows that although the flap undergoes significant motion, the flap itself is rigid. This is relevant, since groups are now targeting flap interactions to establish inhibitory binding specificity. The cantilever motion reported by Beveridge in the dimer simulation is not prevalent in the monomer. The overall protein confirmation is stable on the 160ps time scale. The ends are highly flexible and important for dimerization, consistent with experiment. The x-ray structure of the dimer is a suitable starting structure to block dimerization. This is relevant because groups have now developed nanomolar inhibitors to block dimerization and seek to improve them.

Other applied research on molecules of biomedical interest includes using molecular dynamics simulations to predict function or structures of peptides and proteins. Projects include examination of the role of cyclic AMP in the stability and activity of catabolite gene regulatory protein, a molecular dynamics simulation of the L-alanine crystal, and modeling and simulation of winter flounder ice inhibition peptides.

Basic research is underway to provide a better understanding of biochemical systems. Projects include studies of environmental effects on protein dynamics, a simulation study of interleukin-1b, comparison with crystallographic and NMR structures, harmonic analysis of large systems, modeling and simulation of lipid bilayers, studies of the B-Z junction of DNA, structural analysis of T4 lysozyme mutants in the harmonic limit, cooperativity of cystine bridge formation in peptides and proteins, comparison of simulations on staphylococcal nuclease with NMR data, and structural characterization of a heme:myoglobin adduct.

Findings from the environmental studies of protein dynamics show that the glass transition temperature and the high temperature behavior of globular proteins can accurately reproduce experimental values. There remains a discrepancy for the extent of motion at low temperatures (below 200K). The hydration studies of proteins are consistent with the experimental evidence that only 300 water molecules are required to hydrate a protein with 150 residues. The simulations show that these water molecules are not distributed as a monolayer, but actually cluster to hydrate the charged and polar

groups and to fill in surface pockets, thus leaving large portions of the surface bare.

The 500ps simulation of interleukin-1b provides a molecular picture of the nature of the slow internal motions that have been inferred from nuclear magnetic resonance relaxations experiments. The simulated decay of N-H vector orientational correlation functions suggests that, in addition to fast librational motion, these vectors undergo large jumps in amplitude between conformations stabilized by hydrogen bonds.

Development of Theoretical Methods for Studying Biological Macromolecules

B.R. Brooks, Ph.D.; I. Chandrasekhar, Ph.D.; R.L. Heck; M. Hodoscek, Ph.D.; R.J. Loncharich, Ph.D.; W.M. Southerland, Ph.D.; P.J. Steinbach, Ph.D.

Algorithms and software for new theoretical techniques are being developed and characterized. Ongoing efforts in this area include: new methods for performing free energy perturbation calculation which allow connectivity to change and have a higher accuracy than existing methods, methods for treating solvent implicitly to provide for hydrophobic effects without the explicit inclusion of many water molecules, methods to properly treat electronic polarization in molecular dynamics simulations, and the development of a more accurate flexible water model. Projects include the analysis of hysteresis in free energy perturbation simulations, slow growth homology modeling, development of quantum mechanical potentials and appropriate simulation algorithms for use in molecular dynamics simulations, studies of excited state and electron transfer processes in biological systems, semiempirical Hartree-Fock calculations of proteins, new methods for long-range truncation of the energy potential, and further refinement and examination of free energy techniques.

Many of the parameter sets and models that are generally available are of the quality required for accurate simulation of macromolecular systems. Therefore, parameter development efforts are restricted to areas of primary interest where the existing parameter sets are inadequate. *Ab initio* chemistry, crystal simulations, vibrational analysis, solvated molecular dynamics simulations, and free energy simulations are being used in this effort. One such example is in the development of parameters for simple organic substituents to use in modeling lipids. Projects include development of van der Waals parameters for methylene and methyl groups, development and use of a polarizable and flexible water model, molecular dynamics simulation studies of DNA, analysis of the parameter sets, and conversion of physical models into three-dimensional coordinates for computer analysis and simulation.

Development of Advanced Computer Software

B.R. Brooks, Ph.D.; M.J. Field, Ph.D.; M. Hodoscek, Ph.D.; R.P. Hughey, Ph.D. (U.C. Santa Cruz)

Software and hardware required to obtain optimal performance for simulation research and to develop new tools for future research are being developed. These include microcoding commercial processors and engineering specialized, high-performance computers for specific computational chemistry needs.

Massively parallel high-performance computers hold great promise for the future of high-speed scientific computing. An Intel i860-based machine of this type, acquired, housed and supported by Dr. Robert Martino and associates in CSL, is being used for algorithm development and scientific computing. Methods for this class of machine are being designed and tested, with the goal of using this system for future research where appropriate. A full feature version of CHARMM is being further developed.

Publications

Chandrasekhar I. Parameter development for molecular dynamics simulations of lipids. In: *Biomembrane structure and function: state of the art*, Albany, New York, Adenine Press, 1991 (in press).

Chandrasekhar I, Clore GM, Szabo A, Gronenborn, AM, Brooks BR. A 500 ps molecular dynamics simulation study of interleukin 1beta in water: correlation with nuclear magnetic resonance spectroscopy and crystallography, *J Mol Bio* 1992 (in press).

Janezic D, Brooks BR. Harmonic analysis of large systems. Application to bovine pancreatic trypsin inhibitor, *Periodicum Biologorum*, 1991;93:2.

Loncharich RJ, Brooks BR. Temperature and phase dependence of protein dynamics: a simulation study of myoglobin. In: Renugopalakrishnan V, Carey PR, Smith ICP, Huang SG, Storer AC, eds.) Proteins: structure, dynamics and design. Leiden, The Netherlands: ESCOM *Sci. Publ. B. V.* 1991.

Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-Acetylalanyl-N'-Methylamide Biopolymers 1992 (in press).

Milne GWA, Nicklaus MC, Hodoscek M. Molecular modeling in solvent, *J Am Chem Soc* 1991 (in press).

Raghunathan G, Seetharamulu P, Brooks BR, Guy HR. Models of delta-hemolysin membrane channels and crystal structures, *Proteins* 1990;8:213-225.

Steinbach PJ, Loncharich RJ, Brooks BR. The effects of environment and hydration on protein dynamics: a simulation study of myoglobin: special issue on protein dynamics, *Chem Phys* 1991;158:383-394.

Venable RM, Widmalm G, Brooks BR, Egan W, Pastor RW. Conformational states of a TT mismatch from molecular dynamics simulation of duplex d(CGCGATTCGCG), *Biopolymers* 1992 (in press).

Richard J. Feldmann

Experience over many decades at the NIH and elsewhere has shown that collaboration between scientists who have different interests and functions generally strengthens the activity of each individual. The style of collaboration depends very much on geographic and economic factors. We have been investigating the interaction between computer technology and the style and content of scientific collaborative efforts.

When computer graphics equipment was very expensive, the model that we evolved had the biomedical scientists coming to the central facility to work with computer specialists. When computer equipment in DCRT was unique, scientists were willing to travel from all over the world to use it. In order to be able to solve biomedical problems, it was necessary in the past to spend years programming the computers and the graphic displays. In recent years, as PCs and workstations have become less expensive, more powerful and more widely distributed, we have changed the model for scientific use. A wealth of very powerful programs has appeared in the marketplace making it reasonable for us to purchase and use rather than create applications. The evolution of the campus-wide network has made it possible for scientists to share programs and data while working in their own laboratories.

The connection of the campus network to the NSF-sponsored InterNet has further increased the scope of possible scientific collaborations. We are just beginning to experiment with the potential uses of the InterNet as a scientific computational tool. Each month, more and more sites throughout the world are connected to the InterNet. We are happy when e-mail connections are established to some sites in the third world, while across the United States, Europe and Japan, computational connections are now real. It

becomes possible to think of doing computations across the InterNet which are impossible within one's own laboratory, institution or even country.

The idea of a collaboratory has emerged as a possible way of doing science using the InterNet. Programs and data can be prepared simultaneously and independently by scientists in different geographic places. Both data and programs can be shipped back and forth across the InterNet, or a user at one site can execute a program at another site using data from yet additional sites. As an example, we have formed a collaboratory for doing genetic information processing by bringing together groups from the NIH (DCRT, NLM), from the American national laboratories (Argonne, Lawrence-Berkeley), from academic institutions (Washington University), and from foreign national laboratories (ICOT). In a year the collaboratory has gone from conceptualization to a prototype program running on medium-priced workstations to a deliverable program running on any PC.

With the ICOT laboratory in Tokyo, we have used the collaboratory approach to develop models for studying the folding of proteins. Our Japanese colleagues have written a program that uses a massively parallel logic-programming computer under development in their laboratory. Data, programs and computational results go back and forth across half of the world in seconds. Protein folding simulation using conventional molecular dynamics methods require 10^{10} more computer power than is currently available. Our investigations have focussed on understanding the structural and statistical characteristics of proteins so we can develop simplified models that require less computer power for simulation. In the last year we have looked at several different models at various scales of temporal and spatial resolution.

Collaborations between grants administrators and computer specialists involving the use of computers and networks has also produced important results. Since starting 4 years ago, we have made progress towards moving the NIH grant application process from one that is paper based to one that is completely electronic based. Analysis and simulation of the flow of information through the grant review process showed that substantial gains in efficiency could be achieved by having the applicants prepare the grant proposal in electronic form using their personal computer. Currently, only about 1% of the grant application data is kept in electronic form. Review of all facets of the use of grant application data before and after grant award has revealed which pieces of information are important to a variety of different administrative users. Plans have been advanced for field trials of the grant application software as a precursor to general use.

For the past several years we have been searching for computer graphic tools that could be used by both scientists and administrators. A modular approach to computation and visualization has emerged from discussions several years ago between DCRT and the Steller Computer Corp. The Advanced Visualization System(AVS) presents a possible way of using a variety of different computers across the InterNet to do special-purpose computations, super-computations, database accessing and visualizations. Tedious programming tasks that formerly took weeks can now be accomplished by the now familiar actions of pointing and clicking with an electric mouse. An international repository in North Carolina will make possible the network sharing of many different types of computation and visualization modules.

Whether the global network of computers is used for scientific or administrative work, it is becoming clear that in the next years it will transform the way we do science.

Analytical Biostatistics Section

Peter Munson, Ph.D.

The Analytical Biostatistics Section develops, tests and applies new statistical, mathematical and computational techniques and programs to assist in ongoing studies of molecular biology, physiology, pharmacology, endocrinology, protein structure and related research areas. Staff applies and refines existing theoretical techniques to practical problems in a form easily disseminated and utilized by research biochemists, pharmacologists and clinical investigators.

The section utilizes the theory of mathematical statistics, numerical analysis, and algorithmic and computational methods. Modern, computationally intensive statistical procedures are frequently employed on mainframe, micro and newer distributed, vectorized and parallel architectures. Systems analytic and computer science techniques are used to design, program, document, support and distribute software for specialized problems to investigators in the scientific community.

Major Activities

We have used modern regression techniques applied to subsets of the Brookhaven database of protein structures to discover new nonlinear rules for prediction of protein structural characteristics from peptide residue content and sequence information. Careful cross-validation techniques were used to screen out fortuitous and spurious associations between structure and the sequences of the limited number of (roughly 100) proteins currently available for analysis. Few, if any, prediction rules can exceed multiple linear regression applied to residue proportions, or to selected biochemical residue properties, in ability to predict the total amount of alpha helix within a protein of unknown structure. The best rules can predict three structural classes of proteins (alpha helix, beta sheet, coils) about 55% of the time, overall.

We have attempted to develop new and improved algorithms for the prediction of protein secondary structure from sequence data. Current methods are reported to be about 65% accurate in predicting alpha helix and beta sheet locations. It has been suggested that this figure is limited by the available databank of protein structures, but that further improvements are theoretically possible. We are attempting to find such improvements by incorporating protein class information and by analyzing the patterns of errors (residuals) made by existing methods.

We have created a novel algorithm for aligning multiple protein or DNA sequences that uses much less computation time than previously published methods for the same problem. Optimal alignment of sequences is useful for identifying conserved regions, establishing evolutionary pathways, and finding distantly related sequences. The optimum can easily be found when comparing only two sequences of typical protein length. However, when six or seven sequences are involved, finding the optimum can be computationally prohibitive. Using our iterative randomized approach, the problem is broken down into a sequence of pairwise alignments between subgroups of the original group of sequences. Our method and computer program MUSEQAL, written for the PC, requires at most about seven steps to find the mathematically optimum alignment of five sequences.

We have continued our development and validation of techniques for analysis of pituitary hormone time series. We have developed a new pulsatility index, and applied it to luteinizing hormone (LH) measurements in hemosiderosis patients. The new measure of pulsatility is able to discern a qualitative

change in the patient's time series, compared to an extensive series of normal controls, whereas most other indices (e.g., peak frequency) used previously do not find significant changes. It appears than hemosiderosis patients may have reduced pituitary function, which may explain delayed sexual development.

A mathematical model for the binding of estrogen, its dimeric receptor and DNA was developed in an attempt to explain the apparent positive cooperative binding of estradiol observed by some laboratories and its inhibition by partial agonists and antagonists. While analytical and numeric exploration of this model confirmed that it could produce qualitatively similar binding results, upon further inspection of the available data, it cannot be ruled out that the reported positive cooperativity is an artifactual result caused by measurement and other sources of error.

Other Collaborations

With Dr. Tommaso Costa (LTPB, NICHD), details of the equilibrium ternary complex model for ligand binding to membrane-bound, G-protein-linked receptors were worked out, and a simulation program was written in MATLAB. As the model incorporates the first phenomenological steps by which receptor binding begins the signal transduction pathway, we were able to formalize the notion of antagonism and agonism. A parameter, alpha, of the model indicates the degree to which binding of hormone (H) to receptor (R) also affects the coupling of R to the G-protein (G). The model predicts two phenomena which are observed in some systems: spontaneous G-protein activity in the absence of H, and the existence of "negative" antagonist ligands. Ordinary antagonists manifest their effect by displacing agonist ligand from receptor, but do not propagate a "signal" and so reduce the biological effect associated with binding of agonist. Negative antagonists actually countermand the agonist signal by issuing opposing orders to the cell, even though they bind to the same receptor. The existence of negative antagonists in a receptor system would also imply the presence of spontaneous activity of unbound receptor.

With Dr. Joshua Zimmerberg (LTPB, NICHD) and Dr. Fred Cohen (Dept. of Physiology, Rush Medical College), we applied statistical techniques such as smoothing, filtering, nonparametric density estimation and cross-validation to the analysis of a large collection of records of the lifetimes of exocytotic pores formed between membranes of granules and the cell membrane. This pore formation and growth is the first of a series of events that mediate membrane fusion. Pores are measured indirectly by means of conductance measurements made on a millisecond time scale. In preliminary analyses, we were able to generate an exposure-time-corrected dwell-time histogram, which gives the proportion of the life of a typical pore spent at particular sizes as it grows. Eventually, pores reach a critical size above which they grow explosively until the membranes for granule and cell fuse completely and the granule contents are ejected into the extracellular space. Statistical tests were developed to assess the hypothesis that pores grow in a quantal fashion, and appear to linger at preferred sizes.

With Dr. Enrico Rovati (Institute of Pharmacological Sciences, University of Milan, Italy), and Richard Shrager (LAS, DCRT), we are developing a program to analyze kinetic binding experiments from receptor systems with more than a single class of receptors. The usual experimental design calls for an association experiment (binding followed over time) followed by extensive dilution (dissociation of binding followed over time). The two steps are repeated at varying initial concentrations of ligand. The program is to be written in MATLAB, and requires a numerical solution to potentially stiff differential equations.

With Dr. Mark Zweig (Clinical Pathology, Warren Grant Magnuson Clinical Center, NIH), we are developing a statistical procedure to evaluate the sensitivity of new immunoradiometric methods for assaying low doses of thyroid-stimulating hormone. Pilot studies were run which indicated that the kit manufacturer's suggested calibration procedure was far from optimal, and could be improved dramatically, but still would not reliably measure the very low concentrations needed for a new study.

Statistical Consulting, Teaching

A "Poster on the Macintosh" course was given April 1, 1991. This procedure uses a set of computer tools to create a poster for a scientific meeting directly within the Macintosh computer. Text, charts and graphs are combined within a page layout program, and the final result is printed as a photographic negative at very high resolution. The final poster is obtained by photographic enlargement. The entire procedure, including photography, can be accomplished within 2 weeks, which is much faster than by conventional means.

The section is frequently called upon for statistical advice by scientists within the intramural research programs. Over the last year, the section provided assistance with the use and application of the LIGAND program to over two dozen investigators. Problems such as analysis of repeated measures-type studies and dose-response studies are frequently presented. In one study, intracellular calcium was measured in control and mas oncogene-transfected cells as a function of angiotensin II. Because of the study design a simple analysis was impossible.

Different doses of angiotensin were administered to different batches of cells. Using a three-way analysis of variance, it was possible to establish that the transfection did indeed have an effect upon the angiotensin-induced intracellular calcium. In another study, investigators had used LIGAND to characterize the binding of MHC Class I molecule to antigenic peptide, and wished to interpret the apparent binding affinities in terms of a model which included two states of the MHC. It is known that the ternary complex model can give rise to the observation of two apparent affinity states, so this model was suggested as a basis for further analysis. In a third project, advice was sought concerning the appropriate modeling of drug synergy in the context of chemotherapy. Mathematical models which do display synergy are often generalizations of multiple ligand, bivalent receptor models which have been previously studied in this section. Suggestions were made concerning a possible "model-independent" definition of drug synergy, based on a generalization of cubic splines to two dimensions.

Computer Program Development and Support

Support for a number of previously developed computer programs including LIGAND, ALLFIT, FLEXIFIT, PULSEFIT and DNAFIT continues. The conversion process for program LIGAND to the Macintosh computer has produced a fully working version, including spreadsheet data entry, integrated graphics, and statistics, with menu operation. Over 600 copies of PC and Macintosh programs were distributed during the last year, about half of these being the LIGAND program. Program PULSEFIT has been improved and completely documented, resulting in version 1.3. This program computes the pulsatility index, identifies peaks, and estimates peak heights for hormone time series.

Significance to DCRT and Biomedical Research

This section endeavors to find, develop and apply the best available statistical techniques to a wide variety of problems in biology, chemistry and medicine. In many experimental studies the quality of the final results is constrained by the adequacy and correctness of the statistical analysis and inference used. Complicated study designs, indirect measurement techniques, limited research materials and reagents, small samples, limited length of a time-series (data-sparseness), and even data richness create challenging and difficult problems for statistical analysis as investigators push available technology to its limits. Methodological research performed in this section addresses such problems by creating and optimizing the necessary analytic approaches.

The relevance of the protein structure prediction from sequence becomes more evident as sequence data accumulates at rates several orders of magnitude faster than structural information. Statistics and statistical methodology therefore will play an increasingly important role in deciphering patterns in the growing sequence and structure databases.

Research into computationally intensive statistical methods may provide solutions to the difficult problem of knowing when extrapolation from a fixed set of data is sound. The cross-validation and bootstrap approaches are examples of this technique. DCRT can provide leadership for NIH in the area of computationally intensive statistical methods.

Future Directions

We will continue to investigate methods for secondary protein structure prediction. We will incorporate protein-class information into new nonlinear regression models for prediction. Attempts will be made to incorporate the known nonindependence of structural states for sequences of residues into an appropriate statistical model. We will continue to explore new statistical prediction methods such as Friedman's projection pursuit regression for intelligent dimension-reduction of such high-dimensional problems. We will continue to refine and edit subsets of the Brookhaven database which will serve as an important and rich source of examples for methods development.

We plan to apply the PULSEFIT technique in collaborative studies where qualitative and quantitative distinction of treatment or patient groups is required. One potential collaborator is accumulating extensive studies of women with polycystic ovarian disorder and will make these series available to use for analysis. Another group at NIEHS has expressed interest in using this method in an extensive series of normal children, employing an improved LH assay technique. Another possible extension of the theory would be the calculation of coincidence measures using the techniques of PULSEFIT.

We will continue to investigate the mathematical statistical properties of a number of methods, particularly semi-parametric approaches to data analysis. These efforts are a continuation of those which led to the development of the program FLEXIFIT and of an improved nonrandomness test. While this project has been inactive for the past year, there are some remaining problems to be addressed including calculating the distribution of estimates for smoothing problems where the degree of smoothing is not fixed and choosing optimal smoothing for the curvature test of nonrandomness.

Publications

Berger MP, Munson PJ. A novel randomized iterative strategy for aligning multiple protein sequences, *CABIOS*, 1991;479-484

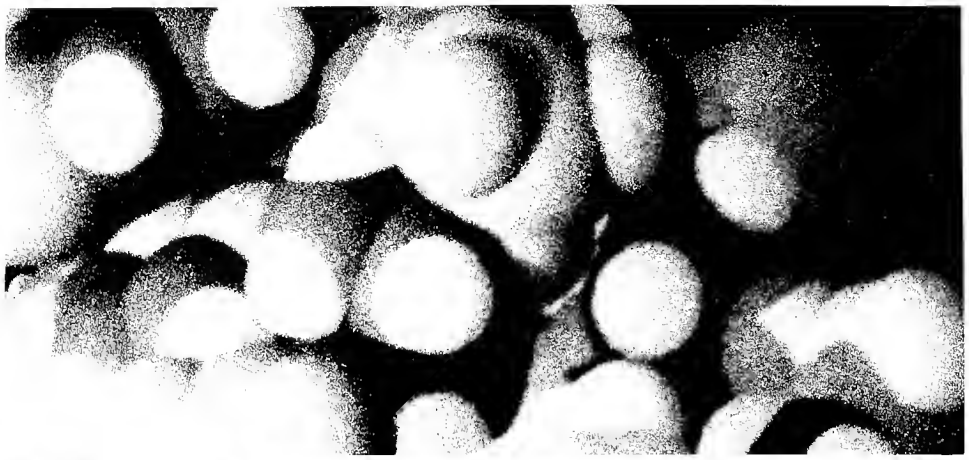
Munson PJ. A review of nonlinear least square curve fitting techniques for data analysis. In: Guardabasso V, Rodbard D, Fotti G, eds. *Computers in Endocrinology: Recent Advances*. New York: Raven Press, 1991,33-44

Munson PJ. Pulsatility index for analysis of hormone time series, presented at joint Statistical Meetings, Atlanta, Georgia, August 18-22, 1991

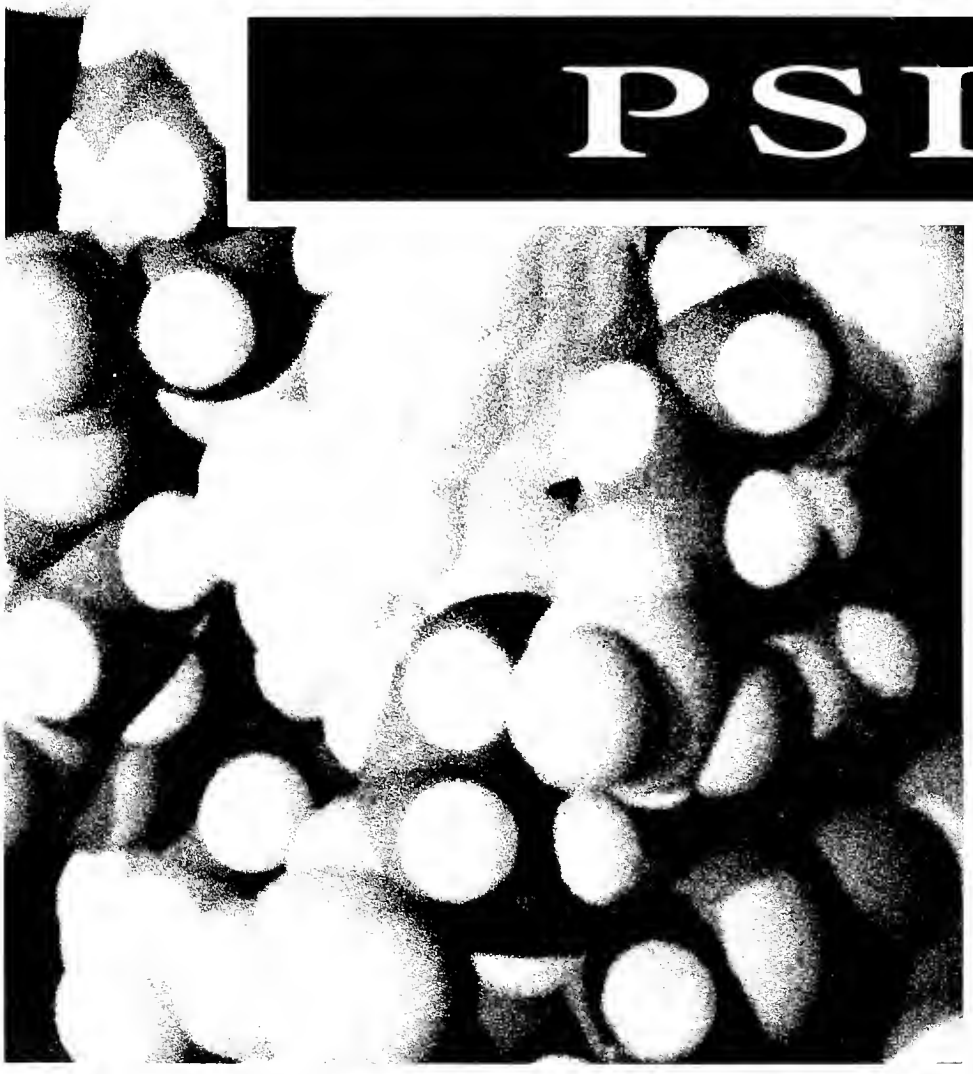
Porrelli RN, Rodbard D. A mass action law model for dimerizing nuclear receptors: effects of partial agonists and antagonists on cooperativity, poster presented at the 73rd Annual Meeting of the Endocrine Society, Washington, DC, June 19, 1991

Rovati GE, Rabin D, Munson PJ. Analysis, design and optimization of ligand binding experiments. In: Maggi M, Geenen V, eds. *Horizons in Endocrinology*. New York: Raven Press, 1991,155-167

Rovati GE, Rodbard D, Munson PJ. Optimization of experimental design for ligand binding studies. In: Guardabasso, V, Fotti G, Rodbard D, eds. *Computers in Endocrinology*: Recent Advances. New York: Raven Press, 1991, 45-58



PSL



Physical Sciences Laboratory

George H. Weiss, Ph.D., Chief

Executive Summary

A considerable effort has been put forth both experimentally and theoretically to develop the use of optical methods for biomedical imaging. The techniques used involve the use of lasers and quasielastic light scattering which are not known to involve risk to humans, in contrast to most methods based on X rays. Additionally, Doppler-scattering techniques can be used to measure parameters that are inaccessable to other imaging techniques, exemplified by the speed at which blood flows in capillaries. In order to translate results from optical imaging techniques into medically useful information a theory must be used that incorporates forward scattering effects into a diffusion picture.

Experiments based on the osmotic stress technique suggest that there are 60 water molecules attaching to the surface of hemoglobin during a switch between the deoxygenated and oxygenated states. The presence of water molecules appears to be necessary for the hemoglobin molecule to open up to accept oxygen molecules. Auxiliary experiments indicate that when the ambient conditions lead to a decrease in the number of bound water molecules, the affinity of hemoglobin to take up oxygen is dramatically decreased.

Accurate approximations have been developed to replace extremely time-consuming computer calculations of the probability density for the crystallographically important three-phase invariant. Methods based on the three-phase invariant are an important component of many widely used programs for producing structural information based on crystallographic data.

Members of the Physical Sciences Laboratory (PSL) perform research and offer consultative services in a variety of areas that emphasize modern concepts and techniques in the physical sciences as they apply to biomedical research at NIH. Some of the areas of expertise of present members of the PSL include applied mathematics, biophysics, chemical physics, crystallography, and statistical physics. The permanent staff is comprised of four permanent scientists, in addition to which there are presently six long-term visitors (i.e., a year or more) and a large number of visitors from all over the world who spend shorter periods of time performing collaborative research with members of the laboratory.

Whenever possible, theoretical work by members of the PSL is combined with experimental work. At present two members of the PSL share laboratory facilities with members of other institutes. Dr. Adrian Parsegian, in addition to being a member of DCRT, is the head of a laboratory in NIDDK in which the researchers are either members of DCRT or NIDDK. The focus of work in this laboratory is on the elucidation and measurement of physical forces that determine biological structure. Dr. Ralph Nossal has joined members of BEIP in setting up a laboratory whose purpose is exploring optical techniques in the biomedical sciences. Present research in this laboratory is aimed at the use of quasi-elastic light scattering to study a number of different properties of tissues.

All of the projects in the PSL rely heavily on sophisticated computer techniques, both for processing experimental data and for the development of relevant theory. A number of new computer algorithms have been developed in the PSL which are used in the wider scientific community.

During this past year, Dr. Parsegian held an additional position: Visiting Faculty with rank of Full Professor in the Department of Physics at Princeton University. One of his aims was to bring to traditional physicists the results of physical force and energy measurements on biological macromolecules. Another purpose, already bringing results, was to establish research links with top physics students interested in

biologically oriented careers. One Princeton graduate student has begun productive work in Dr. Parsegian's NIH laboratory. A graduate Ph.D. in condensed matter physics has also informally joined the group and will be applying for a postdoctoral position. It is Dr. Parsegian's strong impression that a great deal of physics useful to biological research is not being used because of the willful isolation of disciplines. The links forged between this one laboratory and this single department during the past year should strengthen the bio-physical connection for many young people during the next several years.

The work of members of the PSL has been recognized on both the national and international levels, as attested to by the large number of invited seminars given at universities, research institutes, and international meetings. During this past year the PSL organized, conducted and published the papers presented at an international meeting at NIH on models for non-classical reaction rates.

Research Projects

Molecular Forces in Cellular Organization and Function

V.A. Parsegian, Ph.D.
with K. Gawrisch, Ph.D., Sergey Leikin, Ph.D. (DCRT/PSL); D.C. Rau, Ph.D., J.J. Kasianowicz, Ph.D., C. Moore, Marcio Colombo (NIDDK); J.J. Zimmerberg, M.D., Ph.D. (NICHD); E.A. Evans, Ph.D. (University of British Columbia); D.F. Evans, Ph.D. (University of Minnesota); S.M. Gruner, Ph.D. (Princeton University); R. Podgornik (Ljubljana, Yugoslavia); R.P. Rand (Brock University, Canada)

The theme of our work has been to measure, to identify and to formulate the forces that are important to the organization of macromolecules and membranes.

Specifically, using the osmotic stress technique developed in our laboratory, we have measured force vs. distance between membranes or between macromolecules in ordered arrays. We have found that the nature of the forces previously and still commonly assumed in order to explain molecular organization and subcellular assembly are not accurate. The fairly extensive catalog of information that we are developing is allowing us to open new ways to think about the molecular basis of cellular function.

Hemoglobin

Perhaps the salient result of the past year's studies has been the realization that there are 60 water molecules going onto the surface of hemoglobin when it switches between the deoxygenated and oxygenated states. We learned this from seeing that lowering the activity of water (increasing the osmotic pressure of the bathing solution) lowered the number of oxygen molecules that bind to the protein at a given oxygen pressure. Since the osmotic sensitivity of a switch between states should be directly proportional to change in the number of associated water molecules, we can calculate the number of waters per molecule from the difference in oxygen loading curves.

One can now think of water as a ligand much as oxygen and oxygen loading effectors. It appears also that one should review the mechanism of "allosterism" in a large number of proteins, recognizing that the changes in molecular shape necessarily entail energetically significant changes in exposure to water solvent.

The hemoglobin studies have attracted much attention from people interested in developing artificial blood or oxygen carrying systems. The studies have not progressed, however, to the point where a specific system or technique has been developed.

Lipids

From the beginning of their systematic measurement some 15 years ago, no body of force data has been better elaborated than that of interactions between amphiphilic assemblies, particularly between phospholipid bilayer membranes. Progress in this area has been extensively reported earlier by our laboratory. This past year we succeeded in comparing results found with two different techniques: the osmotic stress method and that with the "surface force" apparatus. We examined a synthetic lipid previously reported to show no hydration forces based on measurements with that apparatus. We were able to show that the apparent absence of these forces was due to an artifact of the technique and that osmotic stress measurements provide a more sensitive probe of forces between membranes approaching contact.

We were able also to use force data to test two competing theories of hydration forces, one ascribing it to perturbation of water and another to molecular motion. The data clearly support the former interpretation.

In a theoretical study of hydrophobic force measurements, we created and tested a model of interaction based on desorption and adsorption of solutes. It appears that many measurements of the "hydrophobic" interaction (as opposed to hydration forces) may not be related at all to surface solvation.

Nonbilayer Assemblies

The osmotic stress method has also been used to probe the energetics of nonbilayer assemblies of phospholipids, particularly the inverted HII phase. In this structure, parallel tubes of water, whose axes are arranged on a hexagonal net, are bounded by the polar groups of the phospholipids whose tails fill the intervening space. From the measured work of dehydration under osmotic stress, we have been able to infer the mechanical bending modulus of lipid monolayers. We have also been able to measure the work of creating a phase transition between lamellar and nonlamellar forms. This transition has often been invoked in models of natural cell membrane fusion.

DNA

From the temperature dependence of hydration forces measured between Mn-DNA double helices, we have been able to extract the entropy and the enthalpy of interaction between these molecules. The entropy of bringing Mn-DNA molecules together is positive and appears to be due to the release of ordered water from the vicinity of the macromolecule or the Mn. Just as the total interaction, the entropic and enthalpic parts vary exponentially with distance.

Ionic Channel Proteins

As part of our continuing study of protein conformation and hydration, the alphatoxin transmembrane channel from *S. aureus* has been reconstituted into bilayer membranes and osmotically stressed to provide the information that 60 water molecules enter and leave this channel when it opens and closes under applied voltage. This channel appears also to have different pH sensitivity from each side of the membrane. The large change in the internal volume of water is consonant with results on other channels done earlier in our lab and also with work that has been done in other labs this past year using our osmotic stressing technique.

Biophysical Analysis

R.J. Nossal, Ph.D.
with A. Gandjbakhche, Ph.D., S. Havlin, Ph.D., G. H. Weiss, Ph.D. (DCRT/PSL); R. Bonner, Ph.D., J. Schmitt, Ph.D. (NCRR/BEIP); M. Curran, Ph.D. (NIDDK/LBM); G. Ehrenstein, Ph.D. (NINDS/LB); B. Chance, Ph.D. (University of Pennsylvania); R. Gammon, Ph.D. (Univeristy of Maryland); S. Krueger, Ph.D. (NIST); K. Zaner, M. D., Ph. D. (Boston University Hospital)

For several years we have been investigating models of light propagation in biological tissue. Our motivation has been to provide a framework for implementing optical remote sensing of tissue oxygenation and other physiological properties. Recent studies have involved computer-based Monte-Carlo simulations, carried out to understand how the particular scattering properties of biological tissues influence reflectivities and other measured parameters. Related mathematical expressions have been derived to analyze the surface intensities of transmitted or backscattered light (with R. Bonner and A. Gandjbakhche). Predicted dependences on scattering and absorption cross-sections have been confirmed experimentally, using novel dynamic light scattering methodologies which we have devised to study multiple scattering in turbid colloidal suspensions (with J. Schmitt). Studies also have been performed to develop theories of photon transport in media containing disordered distributions of reflecting inclusions (with S. Havlin and G. H. Weiss).

In addition, we began several collaborations with other NIH researchers who wished to use dynamic light scattering to measure parameters arising in various investigations of cell physiology. Among these are measurements, carried out in conjunction with collaborators' studies of secretion, that are directed towards understanding how neurosecretory vesicles from bovine pituitaries change their volume in solutions of differing osmolarity (with G. Ehrenstein). Another set of measurements relates to studies of structural transformations of the proteoglycan-heparin storage matrix of mast cell granules (with M. Curran), aimed at understanding how a granule matrix condenses to form the dense core of a secretory vesicle to which histamine and other agents bind. In addition to these light scattering studies, correlate measurements have been made at the small angle neutron scattering facility at the National Institutes of Standards and Technology (with S. Krueger), in order to determine the microscopic structure of proteoglycan and other gel matrices (with M. Curran and K. Zaner).

Instrumental Analysis

G. H. Weiss, Ph.D.
with R.J. Nossal (DCRT/PSL); J.A. Ferretti, Ph.D. (NHLBI/IR CH); U. Shmueli, Ph.D. (Tel-Aviv University)

A continuing project in the PSL is the design of optimal NMR experiments in the presence of either instrumental noise or other factors that degrade the performance of NMR instruments. A theory has been developed to estimate the bias incurred in the estimate of peak area by discrete, rather than continuous, sampling. Under realistic circumstances this has been shown to lead to relative errors in area estimates that can be of the order of 100-300%. A project is in the initial stages on the use of NMR techniques to estimate rate constants for some simple chemical reactions. In addition to providing an estimate of the measured parameters, the theory being developed will also allow an estimation of the errors incurred to the particular experimental design.

In past years we have carried out a number of calculations of the probability density for the three-phase

invariant in space group P1, which plays an important role in the application of direct methods of structure determination by x-ray crystallography. These calculations required heavy preliminary mathematical analysis and lead to time-consuming numerical calculations. Results of the calculations have shown that current approximate techniques of the probability can be very badly in error when the crystal is characterized by considerable amounts of atomic heterogeneity as often occurs in the method of isomorphic replacement. We are presently exploring techniques, with considerable success, allowing the experimenter to find accurate approximations to the functions required to translate experimental data into structural information based on our numerical tabulations. Calculations of the probability density of the three-phase invariant are being carried out for other space groups, but these require an enormous amount of mathematical and numerical effort, so that progress is slow at present. A project is underway to estimate the errors in structure determination attributable to different statistical procedures used to compensate for weak reflections. As a first step, experiments are being performed to categorize the type of random noise in crystallographic measurements. Preliminary results indicate that the commonly assumed Poisson distributed noise is not an adequately accurate description of the noise over the range of experiments.

A diffusion and scaling theory has been developed for use in conjunction with light scattering techniques to infer tissue properties from measurements of surface-reflected light intensity from a two-layer medium. The emphasis is on developing simple approximations to the extremely complicated results that come from a more rigorous theory.

Studies in Applied Mathematics and Statistics

G. H. Weiss, Ph.D.
with S. Havlin, Ph.D., J. Bendler, Ph.D. (General Electric); I. Dayan, M. Gitterman, Ph.D., H. Taitelbaum (Bar-Ilan University); R. Kopelman, Ph.D., R. Schoonover (University of Michigan); J. Porra, J. Masoliver, Ph.D. (University of Barcelona); H.E. Stanley, Ph.D. (Boston University)

A major focus of this project has been the development of models for chemical reactions in restricted geometries, as well as models that generalize the commonly used Smoluchowski theory for deriving macroscopic rate laws from a more detailed microscopic picture of the reaction. A number of biologically important reactions occur on cell surfaces. The kinetic behavior of such reactions may not be describable in terms of standard rate equations because they take place in a two-dimensional substrate rather than in a three-dimensional volume. Consequently, one needs to reformulate kinetic equations so as to incorporate microscopic features of the reaction.

A study is being completed of the kinetics of the reaction $A+B \rightarrow C$ in one dimension, in which the A and B are diffusing reactants. The resulting theory concentrates on the behavior of the reaction at early times. Experiments are currently being performed to verify predictions of the theory. An international meeting on non-classical reaction rates was organized and held at NIH. Deviations from classical reaction kinetics can be expected to occur when chemical reactions take place in restricted geometries which preclude perfect mixing of the reactants, exemplified by reactions occurring in cells. The proceedings of the meeting will appear as a special issue of the *Journal of Statistical Physics*. An analysis of the Smoluchowski model that takes many-body effects into account is currently being completed.

A volume in the series "Frontiers in Applied Mathematics" is being edited for the Society of Industrial and Applied Mathematics. The topic of the volume is "Problems in Statistical Physics."

A theory of the diffusion of gasses through polymers has been developed that includes the effects of microscopic heterogeneities in the polymer. The theory is in excellent agreement with experimental data, and may be applied to the diffusion of gases in the lung.

Computer Graphics

B. Lee, Ph.D.
with J. R. Kim, Ph.D. (DCRT/PSL)

GEMM is an interactive molecular graphics package with a friendly and logical user interface that makes it much easier to use than a number of other commercial packages. As in the past years, there have been a large number of inquiries about and requests for copies of the program from outside of NIH. We have significantly enhanced the capabilities of the package over the last year and continue to distribute it free to anyone who asks. The program has been essential for our study of protein folding.

Theoretical Study of Protein Stability

B. Lee, Ph.D.
with J. R. Kim, Ph.D. (DCRT/PSL)

The phenomenon of hydrophobicity can be best understood by combining experimental data and simple, uncomplicated theories not involving many assumptions. Continuing with this approach, it was found in the past year that the large heat capacity change, which is one of the hallmarks of hydrophobicity, is a common response of any liquid upon cavity formation.

It was found recently that this small-molecule behavior can be extended to protein stability. A simple scheme was found that explains the phenomenon of universal temperatures in protein denaturation thermodynamics. The stability of proteins is found to be due to both the hydrophobic and the hydrophophilic parts of the protein, and a quantitative measure was assigned to the relative importance of these two contributions. Future studies will include an accounting for the heat capacity change in protein denaturation. This study is expected to result in being able to construct the whole protein stability curve as a function of temperature and explain the cold as well as the heat denaturation of proteins. It will then give guidance to protein engineers on what should be done to increase the stability of a protein against changes in temperature.

Protein Folding

B. Lee, Ph.D.
with H. S. Kang, Ph.D. (DCRT/PSL)

Our protein folding program, designated "CHORUS," works by using a knowledge base of known structures in a Monte Carlo search procedure. The knowledge base is represented in the form of probability tables for the *phi* and *psi* dihedral angles of the main chain. We found in the last year that this table can predict the regions of the sequence of the protein that act as the folding initiation sites. This is the first time that such sites could be predicted from the sequence alone. The fact that the probability tables can predict such sites demonstrates the power of the information contained in these tables.

During the past year, we tried CHORUS on a number of real systems and found two major problems. The main problem is related to an inadequate potential energy function. A useful potential energy function should be able to discriminate true structure from a myriad of wrong structures. However, no potential function that we have tried so far could do this.

Accordingly, we plan to spend a large part of next year to develop a suitable potential function.

The potential function that has worked best to date was that of Miyazawa and Jernigan. This function is essentially a mean-field potential and represents pairwise contact frequencies that occur in known protein structures. Using this function, we could generate structures that have correct topology but that deviate from the real structure by 6 to 7 Å (in terms of the distance-based root-mean-square deviation). This is about the same level of success that was achieved by workers in other laboratories using a considerably more simplified lattice description of proteins.

The potential that has the best theoretical foundation and therefore the best chance for success is that which combines the hydrogen bond potential and the hydrophobic free energy. Here, we find that the program is highly sensitive to the relative magnitude of these two contributions. The effort to determine the correct relative weight was frustrated by the enormous computer time that is required to compute the hydrophobic free energy. We are currently working on a simplified scheme for computing this free energy and also on using the Intel parallel machine for this purpose.

Computing with an Artificial Neural Network

B. Lee, Ph.D.
with J. R. Kim, Ph.D. (DCRT/PSL)

There is a wealth of information in the database of some 150 protein structures that have been determined by x-ray crystallography and NMR. One of the best ways to extract useful information from such a database is to search for patterns using the artificial neural network (NN) programming technique. In the past year, we wrote a flexible NN programming framework that we could adapt to apply to many different pattern recognition problems. The first problem that we tackled with this approach was the problem of predicting the secondary structure of protein molecules. By using two sets of NNs in two stages, we were able to (1) classify protein structures with over 70% accuracy and (2) to use this information to predict the secondary structures with 74%, 63%, and 63% accuracy for the *alpha*-rich, *beta*-rich and the mixed class proteins, respectively. These numbers indicate that Dr. Kim's program performs better than any that have been reported for this purpose. On the *alpha*-class proteins, in particular, the improvement over the best reported technique is nearly 10%. Dr. Kim is now working on predicting the correct disulfide pairing using the same programming framework.

Publications

Araujo M, Havlin S, Weiss GH, Stanley HE. Diffusion of walkers with persistent velocities, *Phys Rev A* 1991;43:5207-5213.

Ben-Avraham D, Taitelbaum H, Weiss GH. Boundary conditions for a model of photon migration in a turbid medium, *Lasers in the Life Sciences* 1991;4:29-36.

Gandjbakhche A, Bonner R, Nossal R. Scaling relationships for anisotropic random walks, *J Stat Phys* (in press).

Gawrisch K. The energetics of a re-entrant hexagonal-lamellar-hexagonal phase transition sequence in DOPE membranes, *Biochem J* (in press).

Gitterman M, Havlin S, Weiss GH. Effects of an oscillating field on a diffusion process in the presence of a trap, *J Stat Phys* 1991;63:315-322.

Gitterman M, Weiss GH. Some comments on approximations to the cmaster equation, *Physica A* 1991;170:503-510.

Havlin S. Multifractals in diffusion and aggregation. Third Bar-Ilan Conference on Frontiers in Physics. *Physica A* 1990;168:507-515.

Havlin S, Buldyrev SV, Stanley HE, Weiss GH. Probability distribution of the interface width in surface roughening: Analogy with a Lévy flight, *J Phys A* 1991;24:L925-L931.

Havlin S, Kopelman R, Schoonover R, Weiss GH. Diffusive motion in a fractal medium in the presence of a trap, *Phys Rev A* 1991;43:5228-5232.

Havlin S, Nossal R, Trus B, Weiss GH. Photon migration in disordered media, *J Stat Phys* (in press).

Havlin S, Schoonover R, Ben-Avraham D, Kopelman R, Weiss GH. Nearest-neighbor distances in diffusion-controlled reactions modelled by a single mobile trap, *Physica A* 1991;171:232-238.

Juretic D, Lee BK, Trinajstic N, Williams RW. Predicting helices in membrane proteins: A hydrophobic reinforcement method for conformational preferences, (in press).

Kang HS, Lee BK. Separation of the orientational and translational parts of the ring closure problem, *Macromolecules* (in press).

Kang HS, Lee BK. Estimation of protein backbone angle probabilities from protein tertiary structures, *J Mol Biol* (in press).

Kim JR, Lee BK. On even arrangement of points on the surface of a sphere, *Mol Graph* (in press).

Kim JR, Syi SL, Lee BK. GEMM: An interactive molecular graphics package, *Mol Graph* (in press).

Kim JR, Lee BK. Prediction of protein structural class and secondary structure using artificial neural network, *J Mol Biol* (in press).

Kornyshev AA, Leikin S. New approach to the theory of hydration forces, *Biol Membrany* 1990;7:435-442.

Lee BK. Solvent reorganization contribution to the transfer thermodynamics of small non-polar molecules, *Biopolymers* 1991;31:993-1008.

Lee BK, Kang HS. Chorus: A protein tertiary structure prediction program. In: *Recent advances in biochemistry: the Proceedings of 5th FAOB Congress* (in press).

Lee BK. Isoenthalpic and isoentropic temperatures and the thermodynamics of protein denaturation, *Proc Natl Acad Sci* 1991, 88:5154-5158.

Leikin S. On the theory of electrostatic interaction of neutral lipid bilayers separated by thin water film, *J Chem Phys* 1991;95:5224-5229.

Leikin S, Kornyshev AA. Mean-field theory of dehydration transitions, *Phys Rev A* 1991;44:1156-1168.

Leikin S, Kornyshev AA. Theory of hydration forces. Nonlocal electrostatic interaction of neutral surfaces, *J Chem Phys* 1990;92:6890-6898.

Leikin S, Rau DC, Parsegian VA. Measured entropy of water between DNA double helices as a function of separation, *Phys Rev A* (in press).

Masoliver J, Porra WJ, Weiss GH. Solutions of the telegrapher's equation in the presence of traps, *Phys Rev A* 1992;45:2222-2227.

Masoliver J, Weiss GH. Transport equations in chromatography with a finite speed of a signal propagation, *Separat Sci Tech* 1991;26:279-289.

Nossal R, Schmitt JM. Measuring photon pathlengths by quasielastic light scattering in a multiply scattering medium. In: Schmitz KS, ed. Photon correlation spectroscopy: multicomponent systems. Society of Photo-Optical Instrumentation Engineers Proc 1991; 1430:37-47. *Soc Photo Opt Instrum Engr* Bellingham, Washington.

Nossal R, Bonner RF. Differential time-resolved detection of absorbance changes in composite structures. In: Chance B, ed. Time-resolved spectroscopy and imaging of tissues. Society of Photo-Optical Instrumentation Engineers Proc 1991;1431:21-28. *Soc Photo Opt Instrum Engr* Bellingham, Washington.

Oppenheim I, Shuler KE, Weiss GH. Stochastic processes in physics. In: Lerner RG, Trigg GL, eds. *Encyclopedia of Physics*. VCH Publ Inc, New York:1991;1177-1180.

Parsegian VA, Rand RP. Forces governing lipid interaction and rearrangement. In: *Wilschut J, Hoekstra D, eds. Membrane fusion: mechanisms, cell biology, and applications in biotechnology.* Marcel Dekker, New York:1991;65-85.

Parsegian VA, Rand RP. On molecular protrusion as the source of hydration forces, *Langmuir* 1991;7:1299-1301.

Parsegian VA, Rand RP, Fuller NL. Direct osmotic stress measurements of hydration and electrostatic double layer forces between bilayers of double-chained ammonium acetate surfactants, *J Phys Chem* 1991;95:4777-4782.

Pines E, Huppert D, Agmon N. Salt effects on steady-state quantum yields of ultrafast, diffusion-influenced, reversible photoacid dissociation reactions, *J Phys Chem* 1991;95:666-674.

Rabinovich S, Shmueli U, Stein Z, Shashua R, Weiss GH. Exact random-walk models in crystallographic statistics. VI. PDF's of |E| for all plane groups and most space groups, *Acta Crystall* 1991;A47:328-335.

Rabinovich S, Shmueli U, Stein Z, Shashua R, Weiss GH. Exact random-walk models in crystallographic statistics. VII. An all-space group study of the effects of atomic heterogeneity on the PDF's of |E|, *Acta Crystall* 1991;A47:336-340.

Rau DC, Parsegian VA. Direct measurement of temperature-dependent solvation forces between DNA double helices, *Biophys J* (in press).

Rau DC, Parsegian VA. Direct measurement of the intermolecular forces between counterion-condensed DNA double helices: Evidence for long-range attractive hydration forces, *Biophys J* (in press).

Schoonover R, Ben-Avraham D, Havlin S, Kopelman R, Weiss GH. Nearest-neighbor distances in diffusion-controlled reactions modelled by a single mobile trap, *Physica A* 1991;171:232-238.

Taitelbaum H, Havlin S, Kiefer JE, Trus B, Weiss GH. Some properties of the A+B → C reaction-diffusion system with initially separated components, *J Stat Phys* 1991;65:873-891.

Weiss GH. Discussion of "Diffusion in disordered media" by Daniel Ben-Avraham, *Chemometrics and Intell Lab Syst* 1991;10:123-126.

Weiss GH. Moviment browni, difusio, camines aleatoris, *Quaderns Caixa de Pensions* 1990;45:64-71.

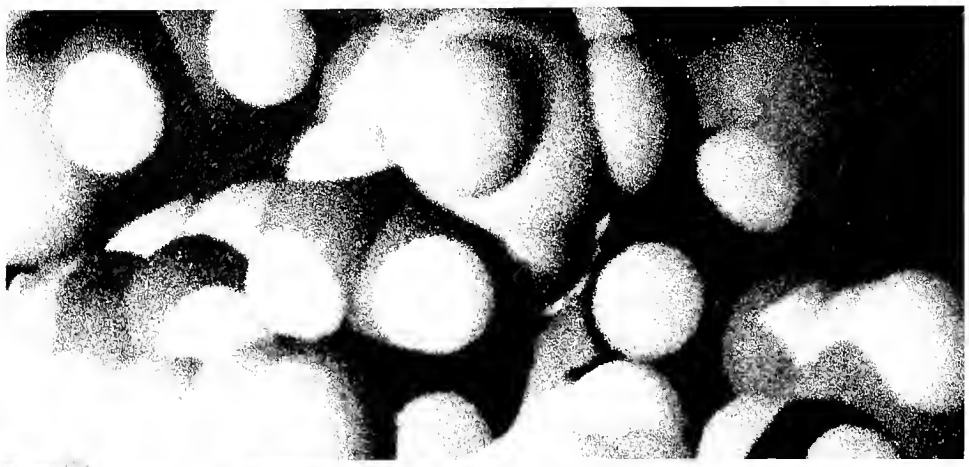
Weiss GH, Gitterman M. Some comments on approximations to the master equation, *Physica A* 1991;170:503-510.

Weiss GH, Havlin S. Some properties of a fractal-time continuous-time random walk in the presence of traps, *J Stat Phys* 1991;63: 1005-1018.

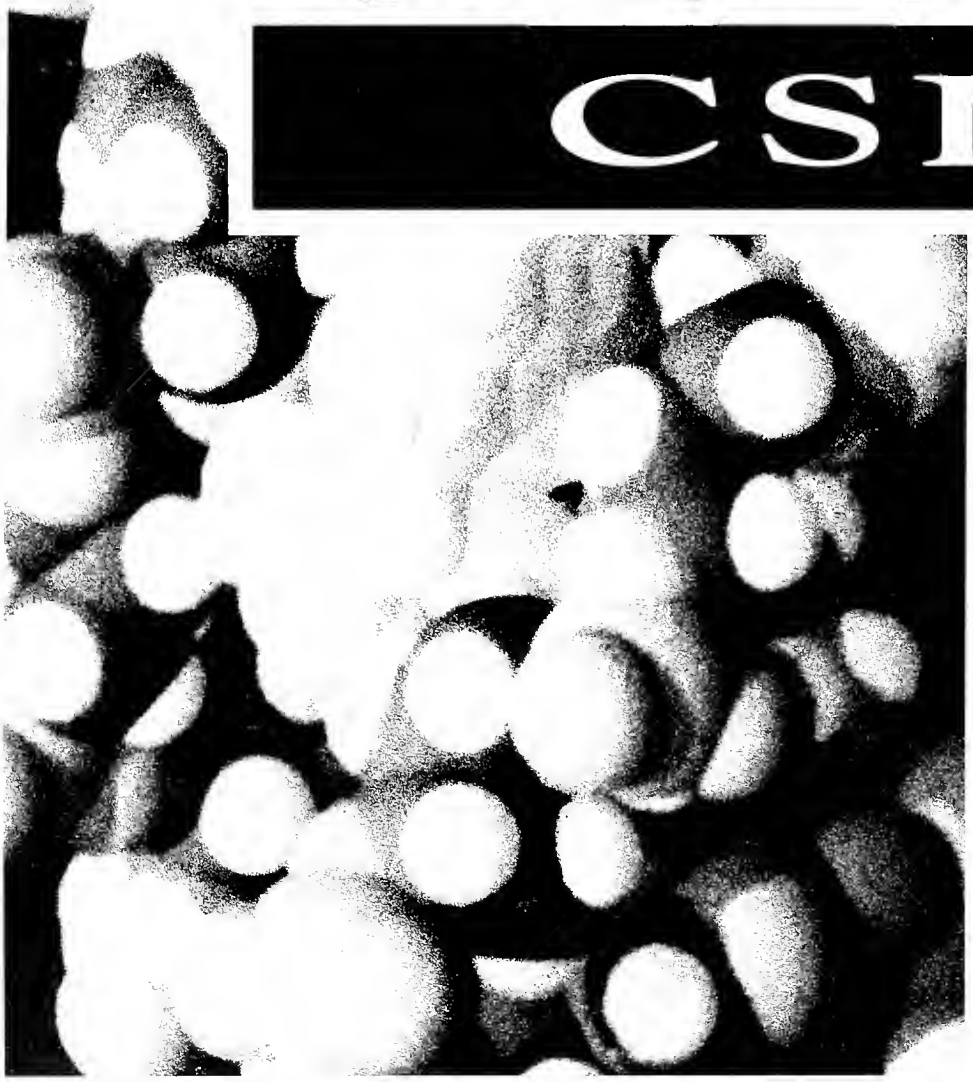
Weiss GH, Kiefer JE, Ferretti JA. Accuracy and precision in the estimation of internuclear distances for structure determination, *J Mag Res* (in press).

Weiss GH, Masoliver J. Nearest trap-particle distances in a one dimensional CTRW model with a mobile trap, *Physica A* 1991;174:209-213.

Weiss GH, Masoliver J. Transport equations in chromatography with a finite speed of signal propagation, *Separat Sci Techn* 1991;26: 279-289.



CSL



Computer Systems Laboratory

Perry S. Plexico, Chief

Executive Summary

The Computer Systems Laboratory (CSL) consults and collaborates with NIH investigators in the development of computer-based systems for laboratory and clinical applications, and conducts computer science and engineering research and development directed toward the application of new computer technologies to NIH programs.

This year, CSL implemented a new experimental computing resource—a 128-processor multiple instruction, multiple data (MIMD) highly parallel supercomputer. Investigation has begun into several important, computationally demanding, biomedical research problem areas, including image processing in structural biology, protein folding prediction, and protein and nucleic acid sequence analysis. Work to speed up the three-dimensional reconstruction of herpes virus images from their electron micrograph projections already has yielded promising results. For example, using 64 processors, the parallel computer is over 500 times faster than the previous sequential system, thus reducing a week's computer time to 20 minutes.

CSL, in concert with the Laboratory Systems Unit, CCB, continued the Advanced Laboratory Workstation (ALW) Project, an effort to introduce wide-scale distributed computing and interoperability to the NIH. The project is oriented to developing a network of powerful Unix workstations for scientists to use in their laboratories, and providing mechanisms for their support via a network distributed file system. In FY91, ALW services first became available to NIH researchers with workstations connected to the campus network. By the end of the year, the project was supporting more than 75 workstations used by scientists from NINDS, DCRT, NIDDK, BEIP, and NIA, among others.

In collaboration with the Clinical Center (CC), CSL is conducting a series of studies that will lead to inclusion of clinical images in the CC's Medical Information System so that images can be electronically transmitted and remotely displayed throughout the Clinical Center. Through this Image Management and Communication System (IMACS) Project, CSL and CC have obtained sheet film digitizers and image communication and display systems. In a pilot project started this year, the Radiology Department uses these systems routinely to transmit chest film images from the film library to the NHLBI Cardiac Surgical Clinic.

A CSL team collaborates with the Receptor Biochemistry and Molecular Biology Section (RBMBS) of NINDS, which engages in large-scale sequencing and gene mapping. This laboratory is attempting to sequence regions of the X chromosome and chromosome 4, and this year began projects to identify and characterize genes expressed in the brain and to sequence the entire smallpox genome. CSL integrates the various computer workstations and network resources and has begun to develop methods to process and archive the extremely large volumes of data that RBMBS generates. The CSL team also has involved the CSL Highly Parallel Computer Project to help improve sequence analysis performance and has integrated the RBMBS computer network with the distributed file system that CSL's Advanced Laboratory Workstation Project implements and supports.

The Computer Systems Laboratory (CSL) conducts research and development in computer engineering and computer science to identify, adapt, and apply state-of-the-art computer technologies to NIH biomedical research programs, and collaborates with scientists and clinicians from other ICDs to devise computer-based systems for specific laboratory or clinical research activities involving real-time data acquisition, computation, or control. Technology areas in which CSL is involved include high-performance computing, laboratory automation, laboratory data processing, distributed computing, computer communications, image processing, and software engineering.

CSL has a staff of 29, plus 5 others who remain on detail to DCRT's Network Task Group. The staff includes engineers, computer scientists, and computer specialists, as well as people trained in the biomedical disciplines who contribute to CSL's multidisciplinary capability and act as an interface to the NIH scientific community-at-large. The staff exhibits the following characteristics:

- familiarity with the rapidly changing computer hardware and software marketplace to enable selection and purchase of appropriate products and technologies
- a thorough understanding of computer engineering, science, and technology to resolve issues associated with developing computer hardware and software, and with integrating individual components into complex systems
- a sufficient background in biology, medicine, chemistry, or physics to communicate with collaborating NIH scientists and to develop biomedical computing applications.

CSL carries on its projects through an Office of the Chief and three sections.

The *Office of the Chief* coordinates planning for new laboratory research and development initiatives, coordinates the work of the various sections to encourage and ensure appropriate cooperation and integration of efforts, and provides overall management of CSL activities.

The *Laboratory and Clinical Systems Section*, under the leadership of Arthur Schultz, collaborates with NIH scientific and medical investigators and applies computer engineering and technology to solve specific laboratory and clinical problems. The section typically undertakes automation projects characterized by real-time interfacing between biomedical instruments and computer system architectures involving workstations and networks.

The *Distributed Systems Section*, headed by Keith Gorlen, investigates the potential of distributed computing technologies, such as distributed file systems, network-transparent graphics systems, and distributed database systems, to aid biomedical research at the NIH. Its principal focus at present deals with networked, interoperable, open-architecture, distributed computing environments.

Dr. Robert Martino heads the *Computational Science and Engineering Section*, which investigates modern, nontraditional, high-performance computer architectures for potential applicability to biomedical problems; conducts image processing as applied to chemistry and biology; operates DCRT's image processing facility; and collaborates with scientists from other ICDs in its use. A member of this section and CSL's principal image processing expert, research chemist Dr. Benes Trus, holds a joint appointment with the Laboratory of Structural Biology Research, NIAMS.

CSL initiates projects when a new computer technology emerges that potentially can benefit the NIH scientific community, such as the highly parallel computer and the advanced laboratory workstation. Other projects help ICDs to solve particular laboratory or clinical computing problems. These projects involve collaboration with a scientist, a group of scientists, an intramural laboratory, or a clinical branch who will use the systems or the methods resulting from the project. Examples include the flow cytometry advanced data analysis project, computer support for high-volume molecular biology sequencing, and investigational work into clinical image communications in the Clinical Center. Finally, CSL offers services that support distributed computing on the campus, and provides advice on computer technology and

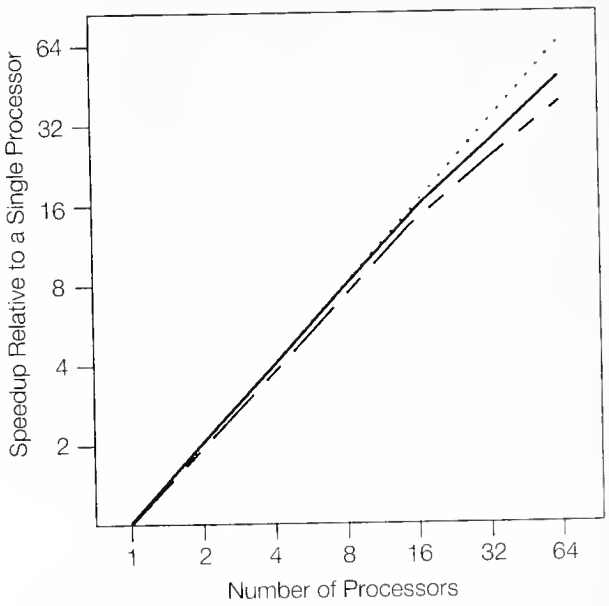
laboratory computing to the intramural scientific staff, and occasionally, to the extramural staff, academic institutions, and other federal agencies. To make maximum use of available resources, CSL emphasizes research and development projects and services that potentially have wide impact on biomedical research at the NIH.

Highly Parallel Supercomputer Speeds Biomedical Calculations

This year, CSL completed implementation of a Touchstone Gamma prototype highly parallel computer obtained through collaboration with the Defense Advanced Research Projects Agency (DARPA). Intel Supercomputer Systems Division developed this multiple instruction, multiple data (MIMD) computer for DARPA. This experimental computing resource now contains a full complement of 128 processors, making it one of the largest and most powerful of such machines in the world. Participation in the Touchstone program has placed CSL in the forefront of research in highly parallel computing.

With system implementation complete, CSL began to investigate applications of highly parallel computing to several important and computationally demanding biomedical research problem areas, including image processing in structural biology, protein folding prediction, and protein and nucleic acid sequence analysis. Promising results already have resulted from work to speed up the three-dimensional reconstruction of herpes virus images from their electron micrograph projections. In electron micrographs, virus particle projections are randomly oriented, and their orientations must be determined in the "find view" step of the three-dimensional reconstruction process. Speeding up the "find view" calculation permits evaluation of more particles, thereby increasing the resolution of the resulting reconstruction image. With 64 processors, the parallel computer is over 500 times faster than the previous sequential system, thus reducing a week's computer time to 20 minutes. Figure 1 illustrates the mechanism underlying this speedup by showing that the "find view" calculation (which determines 3-dimensional orientation of the virus) decomposes almost linearly with the number of processors used.

Figure 1. Speedup of calculations with highly parallel computing.



Key to Figure 1

Ideal Speedup

Find View for a Single Herpes Simplex Virus Particle

Calculation of the Solvent Accessible Surface Area of a 333 Residue Protein

In the area of protein folding prediction, the figure shows similarly good results for the parallel implementation of an algorithm that determines the solvent accessible surface area of a protein. The time required to perform a protein folding simulation can be greatly reduced using the parallel computer as this calculation must be performed at least 200,000 times for a single prediction run. New staff members joining CSL will help to expand parallel computing work in the additional areas of nuclear magnetic resonance spectroscopy, quantum chemical methods, and molecular dynamics simulations.

The President's Office of Science and Technology Policy (OSTP), through the Federal Coordinating Committee for Science, Engineering, and Technology (FCCSET), has initiated a multiagency High Performance Computing and Communications (HPCC) Program to strengthen the Nation's research computing enterprise. During the past year, DCRT participated in the development of an HPCC plan for NIH with the National Library of Medicine, the National Center for Research Resources, and the National Cancer Institute. As a participant in this multiagency program, DCRT will evaluate scalable parallel architectures including associated system software on large biomedical problems, and develop algorithms for these architectures in the areas of structural biology, image processing, and biotechnology.

Advanced Laboratory Workstation Project Goes Production

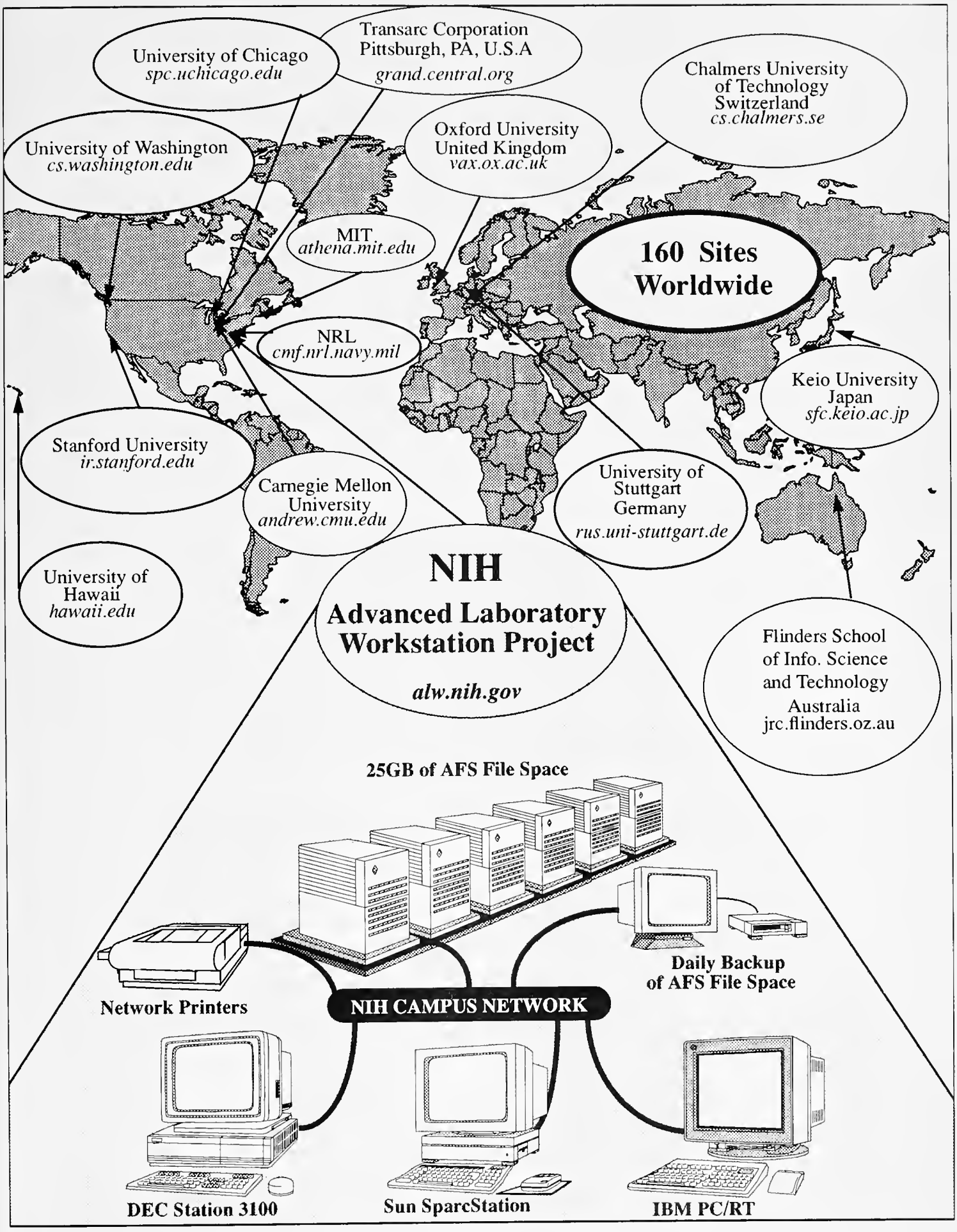
CSL, in concert with the Laboratory Systems Unit, CCB, continued the Advanced Laboratory Workstation (ALW) Project, an effort to develop a network of powerful Unix workstations for scientists to use in their laboratories (Figure 2, next page).

In June 1991, the project passed a major milestone by making ALW services generally available to NIH researchers with workstations connected to the RESnet or NUnet components of the campus network. Investigators in NINDS, DCRT, NIDDK, BEIP, and NIA quickly took advantage, and the number of supported workstations has grown to over 75. Two events spawned this accomplishment. First, we replaced our experimental version of the Andrew File System (AFS) with its first commercial release, making system administration easier and improving performance and reliability. Second, a multiyear contract for AFS fileservers, awarded in March, provided the added disk storage capacity needed to support users.

In preparation for going into production, an ALW field test program began in January, in which ALWs were loaned to investigators for about 6 weeks each. The program interested the investigators in using ALWs, identified and corrected problems, and elicited suggestions for improvements. Twelve groups of researchers in NCI, NLM, NIMH, NINDS, and BEIP participated. This program provided much valuable experience, and produced five new groups of users. We expect to continue the field test program into FY92, as long as significant interest by the NIH research community continues.

The field tests taught that a large diversity of system configurations must be supported. The ALW Project therefore developed two mechanisms to customize ALW computing environments: a facility for tailoring the software configuration that resides on a particular client workstation, and another facility to establish a particular user's operating environment. The latter—which includes such items as the user's preferred "look and feel," application programs to start initially, and the size and placement of their windows

Figure 2. Unix Workstation Network.



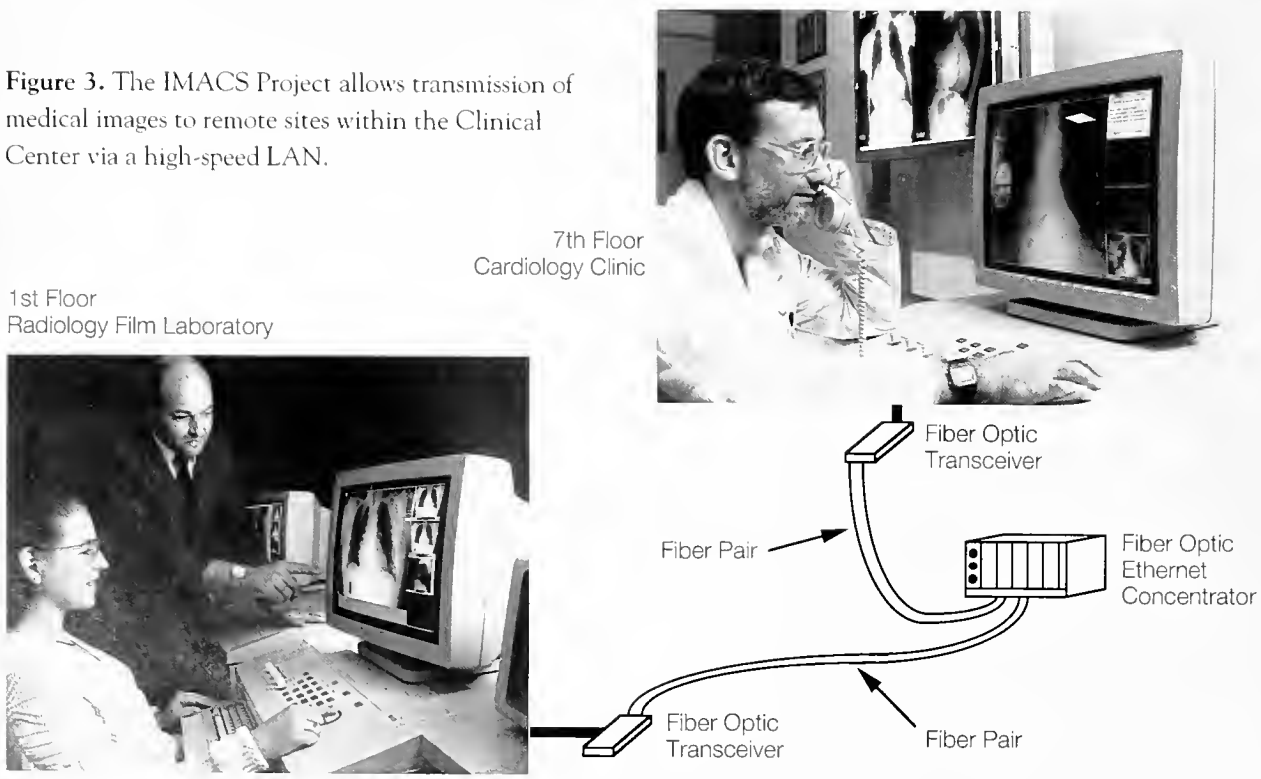
on the display—is particularly important because it is stored in the AFS distributed file system, and thus takes effect at any AFS client workstation the user logs into. This makes the ALW system ideal for use on public Unix workstations, such as those planned for DCRT's proposed Scientific Computing Resource Center: when a user logs into a public workstation, it is dynamically reconfigured to provide the same computing environment he/she normally has when using a workstation in his/her lab.

Medical Images Transmitted Electronically to Outpatient Clinic

In FY90, CSL and the Clinical Center began a collaboration on an Image Management and Communication System (IMACS) project. Various sources of clinical images will be interfaced to the Medical Information System (MIS) via a LAN and an image gateway, so that images can be made available to clinicians throughout the Clinical Center by electronic means. A gray-scale sheet film digitizer that will function as a principal element of this system was obtained.

This year, the IMACS Project obtained two image display systems that are compatible with the film digitizer. They allow transmission of medical images to remote sites within the Clinical Center via a high-speed LAN. A pilot project was begun to transmit chest films routinely from the Radiology Department's film library to the NHLBI Cardiac Surgical Clinic, which meets one morning each week (Figure 3). Members of the medical staff in this outpatient clinic are gaining experience using electronic images in place of film to evaluate radiographic evidence of changes in cardiopulmonary status.

Figure 3. The IMACS Project allows transmission of medical images to remote sites within the Clinical Center via a high-speed LAN.



Kenneth Kempner, a senior member of the CSL technical staff who leads the IMACS effort, also represents NIH with the American College of Radiology-National Electrical Manufacturers Association (ACR-NEMA) working group for medical image communications standards. This group deals with standards for communication between radiology information systems, hospital information systems, and picture archiving and communication systems. Mr. Kempner proposed the concept of applying object-oriented analysis to the ACR-NEMA framework, and arranged for a CSL contractor, Dr. Sanford Orlow, to present a lecture about object-oriented analysis and design techniques to the working group. CSL support for this standards effort will have long-term benefit for both the medical community at-large and the NIH.

Computer Support for Molecular Biology and Genetic Mapping

Molecular biology relies increasingly on computer technology for protein and DNA sequence analysis, management of large databases, and worldwide communications, giving rise to the new term, "molecular biology informatics." In large-scale projects, such as sequencing an entire gene or a large virus, computers have become indispensible. To improve the potential for CSL contributions in this area, James Sullivan and Arthur Schultz designed and coordinated molecular biology training for CSL staff members and other interested DCRT employees. The course was developed and taught by FAES through a contract arrangement, and provided an introduction to important topics in molecular biology to engineers and computer scientists with little previous exposure to biology.

A CSL team led by John Powell continued its collaboration with Dr. Craig Venter's Receptor Biochemistry and Molecular Biology Section (RBMBS) in NINDS. Dr. Venter's laboratory engages in sequencing regions of the X chromosome and chromosome 4, and this year began projects to identify and characterize genes expressed in the brain (with the Department of Energy) and to sequence the entire smallpox genome (with the Centers for Disease Control). CSL contributes to this effort by integrating the various computer workstation and network resources, and has begun to develop methods to process and archive the extremely large volume of data that RBMBS will generate. Specific activities this year include beta testing the Unix version of the Genetic Computer Group's (GCG's) sequence analysis software suite; devising parallel computer approaches with the CSL Highly Parallel Computer Project to improve sequence analysis performance; and integrating the RBMBS computer network, with the Andrew File System (AFS) being implemented and supported as part of the Advanced Laboratory Workstation Project. Because of its security features, AFS is used to manage controlled access to data from the smallpox project by NIH and CDC collaborators, suggesting that AFS could become a principal vehicle for sharing data among geographically distant collaborating laboratories.

Computers to Assist Pharmacists

In a project that demonstrates the varied nature of CSL's collaborative projects, James DeLeo has worked with Clinical Center Pharmacy Department investigators to develop a computer interviewing system that takes medication histories. In the Clinical Center, clinical pharmacists participate in patient care by dispensing pharmacotherapeutic information to patients, assessing patient compliance with their medication regimens, screening for adverse effects caused by single drugs and drug combinations, and sorting out pos-

sible complications between medications, food, allergies, and other medical conditions. Computers can help pharmacists by interviewing patients to take medication histories, then combining patient responses with information from drug databases to detect adverse effects. A computer interviewing system can accomodate many patients, making more pharmacist time available for patients who are not suitable candidates for direct computer interviewing.

The system developed by CSL interviews patients to solicit medication regimen, medical conditions and symptoms, medication compliance, dietary history, occupational and environmental toxic exposure and other information needed to assess adverse drug reactions (ADRs) or interactions. At the end of the automated interview, the system produces a concise report for the clinical pharmacist and the attending physician.

This year, the United States Pharmacopeial (USP) Convention Drug Information database was chosen for use with the interviewing system. Because the database contains adverse clinical effects in lay terminology, it enabled development of a lay-terminology thesaurus to access drug information. USP expects to use the CSL-developed thesaurus to standardize adverse effects terminology and eliminate redundant terms for describing the same effect. CSL also developed a machine-learned ADR detection and evaluation scheme using Delphi polling results obtained from 10 clinical pharmacists that is being integrated into the interviewing system.

Plans for the coming year include refining the thesaurus, developing a method for extracting data from the USP natural language database, completing system integration, and beginning preliminary testing of the integrated system.

VAX/VMS Services

CSL continued its popular VAX/VMS services that began in FY90 as an outgrowth of VAX/VMS-based projects in CSL. With the coordination of CSL's Luther Barden and Dr. Ramon Tate, a contractor-operated "VMS Hotline" supplements consultation provided by CSL staff. The Hotline lets intramural VMS users get help with problems and obtain assistance with common system operations such as software installation or configuration changes. Services are furnished by telephone, by electronic mail, or at the customer's site, whichever is the most appropriate. The level of support service improved this year with the implementation of a VAX cluster that accommodates most of the language compilers, software tools, and network services commonly used at NIH. CSL also organized and sponsored the Biomedical VAX/VMS Users' Group that is devoted to the discussion of topics of interest to NIH VAX/VMS users.

As an active participant in furthering campus networking activities, CSL provides the necessary coordination for attaching new DECnet hosts to the campus network. As a result, DECnet usage of the network has undergone nearly trouble-free growth to more than 175 systems. Because an increasing number of PCs and Macintoshes use DECnet on LANs served by VAX/VMS systems, the number of DECnet nodes is expected to grow to over 300 in the coming fiscal year.

New Technical Services Contracts Supplement Staff

A new, 5-year technical services contract that considerably expands upon the scope and level of effort provided under previous contracts was awarded this year. By supplementing staff resources, it frees CSL staff to

address new research and development initiatives and to expand services. The contractor's services supported the Advanced Laboratory Workstation Project, the Advanced Flow Cytometry Data Analysis Project, the DCRT Image Processing Facility, the VAX/VMS Services activity, and several collaborative laboratory automation projects with NIH scientists. Other contract resources obtained via our interagency agreement with DARPA provided expert programming support services for the highly parallel computer.

Future Plans

During FY92, CSL will continue to expand its work in highly parallel computing. Additional staff and new intramural collaborations will bring new biomedical applications related to NIH strategic initiatives in structural biology. In cooperation with other components of NIH, we will join the federal High Performance Computing Initiative. A planned collaboration with the Institute for Computer Applications in Science and Engineering (ICASE) of the NASA Langley Research Center will give NIH access to their expertise in parallel computing. Finally, we will investigate participation in a consortium of federal agencies and universities being organized by the University of Maryland Institute for Advanced Computer Studies (UMIACS) that will give us access to a single instruction, multiple data (SIMD) stream computer system. Access to the SIMD architecture will enable comparative evaluation with our own MIMD machine for various classes of biomedical problems.

The ALW Project will deploy additional AFS fileservers to accommodate an expected substantial increase in the number of users. Completion of a facility begun this year to archive directories of AFS files to DCRT's IBM System 370 should improve usage of available fileserver disk space and should be attractive to users with large amounts of data, such as images or NMR spectra. CSL also must come to grips with two major issues facing the ALW Project: completing a migration from development to service, and identifying sources of long-term funding for fileservers and other resources and services, once they are truly in a production mode.

CSL will coordinate the implementation of a trans-DCRT imaging technologies program that will include image processing, visualization, communications, and management for both clinical and microscopy applications. Program planning began this year under CSL leadership. Preliminary budget work for FY92 and FY93 was done, the outline of an implementation plan was developed, and an *ad hoc* advisory panel to review the plan was scheduled.

Implementation will await the advisory panel's recommendations, but possible next steps include recruiting a program leader from the NIH scientific community, identifying specific projects for the program, recruiting additional image processing experts, contracting for technical support, and planning for the upgrade of DCRT's image processing facilities.

Research Projects

Highly Parallel Biomedical Computing

R.L. Martino, Ph.D.
with C.A. Johnson, T.K. Yap, E.B. Suh, J.C. Pfeifer, B.L. Trus, Ph.D. (DCRT/CSL); B.K. Lee, Ph.D. (DCRT/PSL); B.R. Brooks, Ph.D, M. Hodoscek, Ph.D. (DCRT/MGSU); S. Erwin (Intel Supercomputer Systems Division); A.C. Steven, Ph.D. (NIAMS/LSBR); A. Bax, Ph.D., L. Kay, Ph.D., F. Delaglio, G. Zhu (NIDDK/LCP); T.R. Kerlavage, Ph.D. (NINDS/LMCN); J. Saltz, M.D., Ph.D. (ICASE/NASA); G. Wiegand, Ph.D. (ISTO/DARPA)

Highly parallel computer architectures provide the only means to attain the computational rates demand-

ed by advanced biomedical computing problems. The physical limits of sequential computer technology are being reached as computational needs in the biomedical sciences grow. The goals of the Highly Parallel Biomedical Computing Project are to determine which parallel architectures are best for the classes of problems that arise in biomedical computing, to develop parallel algorithms for advanced biomedical computing problems, and to provide a highly parallel computer facility that will benefit the NIH staff in their scientific computing needs.

The development of parallel algorithms for advanced biomedical computing problems is essential if biomedical scientists are to take advantage of high-performance parallel architectures. CSL is investigating the following significant parallel computing issues in the context of computationally demanding biomedical applications: partitioning a problem into many parts that can be independently executed on different processors, designing algorithms so that delays of interprocessor communication can be kept to a small fraction of the computation time, designing the parts so that the computing load can be distributed evenly over the available processors, and designing algorithms so that the number of processors is a parameter and the algorithms can be configured dynamically. With its collaborators, CSL is developing and implementing parallel algorithms for the following biomedical application areas: Image Processing of Electron Micrographs in Structural Biology, Protein and Nucleic Acid Sequence Analysis, Nuclear Magnetic Resonance Spectroscopy, X-Ray Crystallography, Protein Folding Prediction, Molecular Dynamics Simulations, Quantum Chemical Methods, and Medical Imaging

During the past year, CSL installed a 128-processor DARPA Touchstone Gamma Prototype Parallel Computer that was constructed by the Intel Supercomputer Systems Division. This iPSC/860 computer is a multiple instruction stream, multiple data stream (MIMD) distributed memory system. With this type of architecture each processor is allowed to execute its own separate program, providing a variety of ways the computer can be used to solve problems. This flexibility gives the programmer a number of decomposition strategies in which to break the application into processes and distribute data to the processors, important for the variety of applications encountered at NIH. The system also includes the Intel Concurrent I/O File System that provides a 10-gigabyte fast access mass storage facility for balancing disk input/output with the computational power of the processor nodes. As a participant in the DARPA Touchstone program, DCRT will evaluate prototype systems that it obtains as a partner with DARPA in the development of scalable parallel architectures and collaborate with other program participants in the sharing of system and application software.

Progress was also made during the past year in application development. The speedup provided by the parallel system in performing the computationally intensive steps in the three-dimensional reconstruction of herpes virus images from two-dimensional projections of the virus obtained from electron micrographs will make it practical to obtain higher resolution images of the virus than was possible with the previously used sequential system. While there are many operations in the reconstruction, it is fortunate that the most time-consuming steps fall naturally into the perfectly parallel category where they can be divided into a set of processes that require little communication with one another. This is demonstrated by the timing results shown in the table for the implementation on the iPSC/860 of the Find View step of the reconstruction. As imaged in the electron micrographs, the particles are randomly oriented and their

orientations must be determined with high precision before the reconstruction can be performed. The Find View step determines the orientational setting of each particle in three-dimensional space, and the table shows that it can be decomposed almost linearly with the number of processors used. With 64 processors, the parallel computer is over 500 times faster than the sequential system in this step, thus reducing a week of computer processing to less than 20 minutes (Table 1).

TABLE 1
Find View Step in Three-Dimensional Reconstruction of Viruses
Timing for a Single Herpes Simplex Virus Particle

Number of i860 Processors	Time Required (Minutes)	Speedup	Efficiency
n	T_n	$S_n=T_1/T_n$	$E_n=S_n/n$
1	7.72	1.00	1.00
2	3.88	1.99	0.99
4	1.96	3.93	0.97
8	0.99	7.78	0.97
16	0.51	15.05	0.94
32	0.28	27.56	0.86
64	0.16	49.15	0.77

Good results have also been obtained in the calculation of the solvent accessible surface area of proteins that is used to predict the conformation of these macromolecules. Table 2 shows the timing results for a single area calculation using an approximation method. A typical protein folding prediction requires at least 200,000 steps of change in state, which means that at least 200,000 of these area calculations must be performed. With 64 nodes, the Intel parallel system was approximately 35 times faster than the IBM 3090 computer.

TABLE 2

Calculating the Solvent Accessible Surface Area of a Protein

Timing for a Single Area Calculation of a 333 Residue Protein

Number of i860 Processors	Time Required (Seconds)	Speedup	Efficiency
n	T_n	$S_n=T_1/T_n$	$E_n=S_n/n$
1	6.93	1.00	1.00
4	1.90	3.65	0.91
8	0.98	7.04	0.88
16	0.53	12.87	0.80
32	0.30	23.04	0.72
64	0.19	35.93	0.56

In the past year, CSL implemented a parallel version of the protein sequence analysis program developed at Harvard University called plsearch. This program generates primary sequence patterns from sets of related protein sequences. Using 64 processors, the parallel version is over 340 times faster than the sequential version on a Sun workstation.

In the coming year, work will continue in the application areas mentioned above with additional staff joining the project. This will include continued work on the three-dimensional reconstruction of large icosahedral viruses, protein folding prediction, and sequence analysis. New work will include the application of maximum entropy processing to NMR spectra for the determination of protein structure, the implementation of the Hartree-Fock approximate solution to the molecular Schroedinger equation, and the implementation of the computationally intensive parts of the CHARMM program used for macromolecular dynamics calculations. We plan a collaboration with the Institute for Computer Applications in Science and Engineering of the NASA Langley Research Center so that NIH can benefit from their expertise in parallel computing. We will also consider a possible collaboration involving alternative single instruction, multiple data (SIMD) parallel architectures with the University of Maryland Institute for Advanced Computer Studies.

Advanced Laboratory Workstation Project

K.E. Gorlen
with P.S. Plexico, R.B. Dew, T. Ghebeles, J.J. Knight, R.M. Malick, E. Persky, J.I. Powell, J.E. Sullivan (DCRT/CSL); J. Dickson, Ph.D., S. Bailey, G. Charrick, S. Fellini, J. Small (CCB); N. Hamsath, S. Orlow, Ph.D., N. Shah (Systex, Inc.)

The Distributed Systems Section (DSS) of the Computer Systems Laboratory and the Laboratory Systems Unit (LSU) of the Computer Center Branch are working jointly on a project to provide support for researchers with high-performance Unix workstations manufactured by a variety of vendors. The worksta-

tions are interconnected by the NIH campus-wide LAN, by which they share resources and access services such as file backup and archiving, software maintenance, applications software, online documentation, nationwide electronic mail and news, computation and database servers, laser printers, and a national distributed file system. Applications for Advanced Laboratory Workstations (ALWs) include molecular graphics and modeling, medical image processing, gel analysis, DNA and protein sequencing and searching, statistical analysis, laboratory data acquisition, and desktop publishing.

Following are the activities of the ALW project this year:

We passed a major milestone this year in June by making ALW services generally available to NIH researchers with Sun workstations connected to the RESnet or NUnet. Researchers in NINDS, DCRT, NIDDK, BEIP, and NIA quickly took advantage of this service, and the number of connected workstations has grown to over 75. This major event was the culmination of several prerequisite accomplishments.

We completed the conversion of our production AFS fileservers and all client workstations (about 30 machines at the time) from AFS 2.0 to AFS 3.0, the first commercial release of AFS. This allowed us to support Sun-4 (SPARC) workstations and resulted in improved reliability and higher performance, and made the system easier to administer.

After a 2-year procurement process, we awarded a 5-year, $1.9M contract to a vendor to supply AFS fileservers and maintenance. The new Sun Microsystems SPARCserver 490 fileservers provide much high-performance, sorely needed disk storage: by the end of FY91 our production system, termed a "cell" in AFS parlance, will have expanded to over 25GB, and our test cell will contain 6GB.

In preparation for going into production, we began an ALW field test program. The purposes of this program are to identify and correct system problems, elicit suggestions for improvements from NIH researchers, and to interest them in using ALWs. Since then, we have loaned ALWs for periods of about 6 weeks to over 12 groups of researchers in NCI, NIMH, NINDS, and BEIP. This program has provided much valuable experience, and has thus far gained us at least five new groups of users. We expect to continue the field test program into FY92.

One of the lessons learned from the ALW field tests is that we must be prepared to support a large diversity of system configurations. We have therefore developed two mechanisms for customizing ALW computing environments: a facility for tailoring the software configuration that resides on a particular client workstation, and another facility to establish a particular user's operating environment. The latter—which includes such items as the user's preferred "look and feel," application programs to start initially, and the size and placement of their windows on the display—is stored in the AFS distributed file system and takes effect at any AFS client workstation the user logs into. This makes the ALW system ideal for use on public workstations, such as those planned for DCRT's Scientific Computing Resource Center: when a user logs into a public workstation, it is dynamically reconfigured to provide the same computing environment the user normally has when using a workstation in his/her lab.

The ALW Project encourages users to operate their workstations as so-called dataless clients: no information that cannot be easily reconstructed is stored on the local disk, thus eliminating the need to back it up. Applying this concept has enabled us on several occasions this past year to rescue users who

have experienced disk crashes.

Developing and maintaining the configuration files necessary for managing dataless clients requires a great deal of effort by highly skilled people—this activity currently consumes more staff manpower than any other single activity. Recognizing this, we undertook the development of software tools to help automate the process.

Our major procurement activity in FY91 was trading in our 11 aging Sun-3 workstations for Sun SPARCstation-2 color workstations (we expect delivery by January 1992) and upgrading our two Sun-3 fileservers to Sun-4s so we can run the same software on all our fileserver machines.

The first practical version of the NFS/AFS Translator was included with AFS 3.1, which Transarc (the AFS vendor) released in March, and we distributed to all client workstations in April. We implemented several commands to simplify use of the translator, and have tested it successfully on the Convex, but have found problems in using it with SGI workstations. We are working with Transarc to resolve these.

We installed a variety of new applications and updated most of our old ones in FY91.

This year we began collaborating with the National Institute on Aging on an image processing project which involves using ALWs to perform two tasks: (1) registration of positron emission tomography (PET) and magnetic resonance imaging (MRI) three-dimensional images for the analysis of functional anatomy, and (2) rescaling of MRI images to the dimensions of a standard brain atlas. NIA purchased a SunSPARCstation-2, and we installed AFS 3.1 to give them access to the wealth of programs and file space on our distributed file system. We also installed the Analyze image processing software from the Mayo Clinic and the MATLAB mathematics package. Specialized software developed at Hammersmith Hospital in London was installed, which uses Analyze and MATLAB to process PET images. To facilitate transfer of images from the Nuclear Medicine VAX computer, where PET and MRI images are usually stored, we installed DECnet on the SPARCstation-2.

Work continued this year on the NIDDK Pilot Project started in 1987. We have installed AFS 3.1 on 10 of their Unix workstations, allowing researchers to use our distributed file system for software development.

Previously, most ALW users had been logging into either the Convex or our domain name server host to read their mail. This was inconvenient, and didn't scale well with an increasing number of users. To alleviate this problem, we deployed the Andrew Message System (AMS) and the Andrew Messages Delivery System (AMDS) in April. AMS handles multi-media messages and provides message folders and bulletin boards. AMDS utilizes AFS to deliver mail directly to a user's home directory where it is backed up daily, provides a "white pages" directory service, and employs multiple post office machines for high availability.

The USENET news services became available to SGI workstations this year when we installed software on them to communicate with our news server machine via the Network News Transfer Protocol (NNTP). We also upgraded the news server software to a more efficient version that can store news articles in AFS, where they are automatically backed up regularly and can be more easily managed.

Our training activities in FY91 included contracting with Transarc to offer their AFS User's course, developing and presenting a seminar to introduce researchers to ALW technology, and presenting seminars on object-oriented programming and the C++

programming language. We also began working on an ALW User's Guide, but much remains to be done.

We have started to implement a facility for archiving directories of Unix files to DCRT's IBM 370. This should help us make more efficient use of the disk space on our new AFS fileservers, and should be attractive to researchers collecting large amounts of data, such as images or NMR spectra.

Plans for FY92

In the coming year we plan to expand the ALW user base, most notably by supporting NIH's Multimodality Radiological Image Processing System.

Fiscal Year 1992 will see the use of the NIA image processing system for the registration of PET and MRI images by clinical investigators. The second part of this project will begin with the installation of specialized software developed at the Mitre Corporation for the rescaling of MRI images to a standard brain atlas. A PIXAR computer will perform the rescaling, while the Sun SPARCstation-2 will display the rescaled images.

Early in FY92, we anticipate that the Open Software Foundation will announce its choice for a Distributed Management Environment (DME) technology. We will track this technology closely for potential use on the project, particularly in the area of distributed printing management.

We will attempt to support the IBM RS/6000 and the HP 700 series of workstations.

DCRT also must come to grips with two major issues facing the ALW Project: completing a migration from development to service, and identifying sources of long-term funding for fileservers and other resources and services.

Computer Support for Molecular Biology Sequencing and Genetic Mapping

J.I. Powell, J.C. Kelley
with A. Kerlavage, Ph.D., M. Dubnick, Ph.D. (NINDS/RBMBS)

Molecular biology relies increasingly on computer technology for protein and DNA sequence analysis, management of large databases, and worldwide communications, giving rise to the new term molecular biology informatics. In large-scale projects, such as sequencing an entire gene or a large virus, computers have become indispensible.

CSL continued its collaboration with Dr. Craig Venter's Receptor Biochemistry and Molecular Biology Section (RBMBS) in NINDS. Dr. Venter's laboratory engages in sequencing regions of the X chromosome and chromosome 4, and this year began projects to identify and characterize genes expressed in the brain (the cDNA project with the Department of Energy) and to sequence the entire smallpox virus (with the Centers for Disease Control).

The goal of the cDNA project is to identify and characterize all of the approximately 30,000 genes expressed in the human brain. This project will receive funding from the Department of Energy for 2 years starting January 1, 1992. It will involve sequencing cDNAs from humans (48 templates/day), *C. elegans*, and *Drosophila* (mix totaling 48 templates/day) to produce expressed sequence tags (ESTs). ESTs have applications to discovering new human genes, to mapping the human genome, and to identifying coding regions in genomic sequence. The smallpox project is being done in collaboration with CDC. Its objective is to completely sequence the smallpox virus prior to the destruction of existing virus specimens. The smallpox DNA sequence is about 170 kilobases long. Running about 48 templates/day, it is anticipat-

ed that this project will take about 6 months. However, it is the most demanding sequencing and assembly project undertaken by the RBMBS laboratory, so it will put the automated sequencers and computer software to a severe test.

CSL contributes to these efforts by integrating the various computer-workstation, software, and network resources, and has begun to develop methods to process and archive the extremely large volume of data that RBMBS will generate. Specific activities this year include beta testing the Unix version of the Genetic Computer Group's (GCG's) sequence analysis software suite; evaluating a number of other sequence analysis software packages; installing a new fileserver; and installing the Sybase database management system.

The cDNA project has produced increased demand for sequence database searching. To address this problem, the network BLAST program from NLM was installed, allowing BLAST search requests to be transparently sent to an NLM high-speed remote BLAST server. BLAST also was installed on local Sun workstations to accommodate BLAST searches of the local EST database. A program to search a database of protein library motifs was evaluated on Sun's and on the DCRT Convex computer system. It appears to be a valuable tool, but too slow (90-120 minutes/EST) for the rate of EST generation (approximately 100/day). Early results from a new collaboration with the CSL parallel computer group suggest that the Intel i860 hypercube machine may yield a performance improvement factor of up to 100.

This year, the RBMBS computer network was integrated with the Andrew File System (AFS), being implemented and supported by DCRT as part of the Advanced Laboratory Workstation (ALW) Project. Because of its security features, AFS is used to manage controlled access to data from the smallpox project by NIH and CDC collaborators, suggesting that AFS could become a principal vehicle for sharing data among geographically distant collaborating laboratories.

Dr. Venter's laboratory works at the forefront of DNA sequencing. Changes in emphasis in the past year have shifted the focus from sequence assembly problems to sequence analysis and data archival/retrieval problems. During the coming year, CSL will concentrate on methods to process and archive extremely large volumes of sequence data, and help to make the technologies developed through this project available to other investigators at NIH. For example, we plan to obtain a "fast data finder" system, and make it available as a network-sharable resource for high-speed sequence searching for use by scientists throughout NIH.

Molecular Graphics, Computer Modeling, and Sequence Analysis

B.L. Trus, Ph.D.
with S. Havlin, Ph.D. (DCRT/PSL); A.C. Steven, Ph.D., J. Conway, Ph.D. (NIAMS/LSBR); M. Unser, Ph.D., A. Aldroubi, Ph.D. (NCRR/BEIP)

The purpose of using molecular graphics, computer modeling, and sequence analysis is to gain insight into macromolecular or biological structures.Using molecular graphics, scientists can computationally construct models which may be useful in deciding between two or more alternative interpretations of biochemical or structural data. Computer modeling is often important in understanding biophysics or other biochemical relationships, and how these relate to biological structures. Sequence analysis uses the one-dimensional amino acid sequence of proteins together with both Fourier analysis and other predictive algorithms to attempt to identify parts of the sequence which may have a regular structure.

These interrelated computational methods are used to extrapolate known structural information to predict useful three-dimensional relationships. Often, three-dimensional structural information is unavailable or experimentally intractable. Three studies currently in progress include collaborations with PSL, DCRT to study computer models of biopolymers; with LSBR, NIAMS to predict the structure of macromolecules; and with BEIP, NCRR to create new software for the analysis of gel electrophoresis.

Progress this year has involved continuation of a long-term study of the structure and assembly of intermediate filaments, which has resulted in a new publication. A new study involving computer models of biopolymers has yielded information on mechanisms for diffusion, and has been submitted for publication. Work on understanding the physics of gel electrophoresis has resulted in another publication.

Work will continue on the three phases of this project, with increased emphasis on correlating the secondary structure as determined from the amino acid sequence with the three-dimensional image analysis of electron micrographs (see "Structural Biology: Image Processing of Electron Micrographs" following). The ultimate goal of the use of these methods in structural biology is not only to elucidate the three-dimensional structure, but to be able to draw conclusions about the activity or function of the biological macromolecule being studied.

Structural Biology: Image Processing of Electron Micrographs

B.L. Trus, Ph.D.
with A.C. *Steven, Ph.D., E. Kocsis, Ph.D., F. Booy, Ph.D., J. Conway, Ph.D.(NIAMS/LSBR);* M. *Unser, Ph.D., A. Aldroubi,* Ph.D. (NCRR/BEIP)

This project uses image processing techniques to analyze electron micrographs. In order to answer important questions in structural biology, it is necessary to obtain relatively high-resolution two- and three-dimensional structural information about biological macromolecules. While atomic or near atomic resolution information has traditionally been available by x-ray crystallography for some small molecules and small proteins, the overwhelming majority of biological macromolecules are not crystalline, or are too large and therefore not amenable to 3-D crystallography.

Biological specimens can, on the other hand, be visualized in the electron microscope using a number of specimen preparation techniques. Cryo-electron microscopy, for example, attempts to preserve "native" structure by surrounding the specimen with a layer of ice. Collaborative research with LSBR, NIAMS is currently under way on a number of projects whereby the electron micrograph images are computationally corrected, combined, averaged, reconstructed, or in some way computationally enhanced to improve the signal-to-noise ratio or increase the interpretability of the structures being visualized.

Of particular interest to our research is the understanding of viral structures. Previous efforts have focused on specimens such as bacteriophage T7, beet necrotic yellow vein virus, vesicular stomatitis virus, and some glycoproteins of HIV. At present, we are involved in the investigation of the structure of a large animal virus, the human herpes simplex virus (Type 1). We are in the process of determining the location of the major capsid proteins. Using the three-dimensional icosahedral reconstruction technique, we apply the symmetry of these virus particles both to find the orientation of randomly oriented particles (in ice) and to combine many particles into a three-dimensional reconstruction. Biological material for these herpes reconstructions is provided through collaboration

with researchers at the University of Virginia, Charlottesville. The electron microscopy is performed in LSBR, NIAMS. Interpretation of our 3-D reconstructions is performed jointly by all collaborators.

The precursor herpes nucleocapsid (B-caps) can be made to form empty (A-caps) or full (C-caps) DNA containing capsids. Computer reconstructions, together with supporting biochemical data including gel electrophoresis, are used to correlate the 3-D reconstruction with an interpretation of the changes or differences in the herpes nucleocapsid. In addition, the empty shell is reprojected in the same orientation as the original C-caps images and then subtracted to yield the first clear computer images of the DNA inside a nucleocapsid. Methods such as these are important to understanding the packaging of DNA in large animal viruses. The A-caps and C-caps result has recently been published.

Treatment of B-caps with 2 M guanidine hydrochloride causes the penton capsomers (located on the five-fold symmetry axes) to dissociate, and some of the triplexes (structures which join three capsomers for added stability) are also removed. Removal of the guanidine hydrochloride causes the pentons to begin to reform. A monoclonal antibody to VP5 (the major capsid protein) has been produced, and the 3-D reconstruction shows that the hexons are composed of VP5. These results are being prepared for publication.

A new series of antibodies are also being evaluated to answer remaining questions about the localization of the major capsid proteins. Once the location and function of the major capsid proteins are found, there are some lesser proteins which may be useful to study to complete the analysis of herpes. In addition, we are currently evaluating other icosahedral viruses for potential 3-D reconstruction candidates. Often, unexpected problems arise in these types of analyses. As a result, new computer algorithms are constantly being devised to answer questions in the reconstruction. For example, we often wish to combine data from more than one micrograph. While these data may be nominally at the same magnification, due to differences in defocus, combining different micrographs can, in fact, result in errors of scaling of a few percent. We are collaborating with BEIP, NCRR and LSBR, NIAMS to develop a new method of scaling micrographs so that multiple micrographs may be combined to include more data in each analysis or reconstruction; this work is being prepared for subsequent publication.

Image Management and Communication System (IMACS)

K.M. Kempner
with H.G. Ostrow (DCRT/NTG); T.L. Lewis, M.D. (CC/DIR); J.L. Foy, M.D., Ph.D.(CC/IS); E.E. Tucker, M.D. (NHLBI/CB); G.P. McMahon (CC/DIR); J.F. Fessler (NCRR/BEIP)

Medical images are an important component of the medical record generated during a patient's hospital stay or clinic visit. Unfortunately, these images represent a difficult-to-manage data source due to the tremendous size of the datasets involved. The Clinical Center (CC), like most university and research hospitals, is attempting to solve the problem of how to consolidate medical images with the conventional alphanumeric data contained within its Medical Information System, in order to more completely realize their goal of a comprehensive electronic medical record.

Toward this end, DCRT and the CC are collaborating to develop a series of demonstration projects aimed at supporting image integration into the electronic medical record. Images that concern us range

in size from diagnostic electrocardiograms (up to 16 Kbytes) through tomographic scans (up to 256 Kbytes) to conventional film (nondigital) X rays (up to 4 Mbytes).

Standard 12-lead diagnostic electrocardiograms are automatically acquired, interpreted, and stored on magnetic disks utilizing a Hewlett Packard ECG Management System located within the Clinical Center. In order to transfer ECG waveforms and their related diagnoses from this minicomputer system to the CC Medical Information System, a remote ECG workstation is to be utilized as a gateway between the two systems using a RS-232 pathway. Because ECG waveforms are essentially a binary image (black waveforms on white background) and because the number of equivalent black pixels in such an image is extremely low (approximately 0.1 percent), the ECG waveform data is more efficiently stored and transmitted as a time-ordered list of 10-bit ECG amplitudes, rather than as a 2.75K X 3K pixel image.

Chest X rays are routinely obtained within the Diagnostic Radiology Department, and these images are appropriate for integration into the TDS Medical Information System as well as for transmission to the relevant ACRF clinic where the patient may be seen. A sheet-film digitizer interfaced to the image gateway would handle this function, and we obtained a Vision Ten gray-scale digitizer as an integral part of the proposed image gateway.

We also obtained two Vision Ten display systems, compatible with the sheet-film digitizer. Utilizing the fiber optic network installed within the Clinical Center by the DCRT Network Task Group, communication of medical images between the Radiology Department's Film Library and remote sites became possible.

To gain experience with electronic image transmission, the weekly NHLBI Cardiac Surgical Clinic became the first outpatient clinic to routinely use chest films transmitted over this Ethernet pathway. Within the clinic, electronically transmitted images can be compared with the original film-based image to verify the preservation of fine details in the electronic image. Initial evaluation has shown that detail within the pulmonary vascular beds is the most difficult to visualize, frequently requiring software adjustment of brightness and contrast under the viewer's control.

Brain Image Registration

K.M. Kempner
with M.V. Green, Y.C. Yan, S.D. Stein (CC/NM); C. Kertzman, Ph.D., T.A. Zeffiro, Ph.D. (NINDS/MN); J.F. Fessler (NCRR/BEIP)

An elusive problem faces researchers involved in the correlation of brain form (structure) from x-ray computed tomography (CT) images and brain function (metabolism) from nuclear medicine positron emission tomography (PET) images. The difficulty concerns the superposition and registration of the tomographic views obtained from these two imaging modalities.

The driving force behind the goal of brain image registration is the need to develop a greater understanding of the processes underlying the generation of PET images. It is hoped that development of techniques for the accurate correlation of CT structural data with PET metabolic information will enhance this understanding.

Our approach to this problem is based upon a three-stage solution. First, we are developing practical methods for the accurate and reproducible placement of the head within a tomographic scanner's aperture. Second, we are developing techniques for monitoring head position during the image acquisi-

tion process, so that corrections may be made for head movements before the image is generated. Third, we are developing simplified algorithms for the scaling and registration of digitized images from different scanners, on a digital display subsystem.

Precise orientation of the subject's skull within the scanner's aperture is monitored and recorded through the use of a Polhemus Navigation Systems positron/orientation transduction subsystem connected to an IBM PC-XT.

The development of two inexpensive custom-molded oral appliances allows the position/orientation subsystem's sensor to be fixed to the subject's skull. A novel targeting algorithm was derived to provide the technician operating the system with visual cues related to head position within a scanner's imaging volume.

The Polhemus Navigation Systems position/orientation subsystem allows two independent sensors to be utilized simultaneously. Two-sensor software was completed, and extensive evaluation has begun prior to its experimental use with test subjects.

An additional position/orientation measurement subsystem was obtained by the Nuclear Medicine Department from Ascension Technology. This system is currently being evaluated for linearity, and for sensitivity to nearby metallic objects, a problem common to all electromagnetic-based tracking systems. It is claimed that this device's utilization of quasi-static fields increases its immunity to certain types of metal in close proximity. This has yet to be verified.

Future efforts will center on clinical testing of the accuracy and repeatability of skull placement in tomographic scanners, as well as in refinement of algorithms for removing motion artifacts during the scanning process.

Diagnostic Electrocardiographic System

K.M. Kempner
with E.E. Tucker, M.D. (NHLBI/CB); and J.F. Fessler (NCRR/BEIP)

The Clinical Center's heart station utilizes a computerized Electrocardiogram (ECG) Data Management System to process 12 lead ECGs from all patients within the facility. This Hewlett Packard system:

- collects ECG waveforms
- identifies features
- measures amplitudes, durations and intervals
- provides a clinical diagnosis
- allows editing of diagnosis after physician overread
- stores all ECG waveform data and diagnostic reports.

In addition, the database can be searched to identify patients who fit specific search criteria based on specific diagnosis and/or measurement values.

The medical diagnostic criteria are encoded as If-Then production rules in a diagnostic criteria set. These rules were written using Hewlett Packard's Electrocardiogram Criteria Language (ECL), and the criteria set may be modified by the user to tune existing criteria or add new criteria. This year, Hewlett Packard's newest adult and pediatric criteria sets were implemented on the system, and customized to heart station standards.

ECGs are transmitted to the Data Management System over dial-up telephone lines within the Clinical Center. Originally, the transmission was mediated utilizing analog FM technology; however, this year ECG transmission was converted to an entirely digital pathway. Analog FM modems in the ECG carts and within the computer site were replaced with 2400-baud digital modems.

The heart station initially utilized three ECG machines to obtain clinical ECGs from both inpa-

tients and outpatients. An additional ECG machine is located in the Occupational Medical Services Department. This year, six new ECG machines were obtained, and a large number of existing nontransmitting ECG machines were upgraded into computer-compatible ECG machines. The end result, after all upgrades are complete, will be a total of 40 ECG machines compatible with the ECG Data Management System.

Ultimately, the ECG waveforms and diagnostic reports will be sent to the Clinical Center's Medical Information System (MIS) for display at any of that system's user terminals. A Hewlett Packard ECG Workstation has been obtained for use as the interface between the ECG Data Management System and the Clinical Center's MIS. Communication will occur via an RS-232 link between the two systems. Implementation of this bi-directional pathway is not expected until next year.

High-Resolution Imaging System

M.A. Vivino, A.R. Schultz

CSL has assembled a high-resolution, wide-dynamic-range CCD camera imaging system and has made it available for use by NIH researchers. A Photometrics CCD camera, Macintosh IIfx computer and a Gordon Instruments lightbox are available for high-resolution imaging. The principal component of the system is the camera that, with a thermoelectrically cooled Kodak CCD image sensor, can capture images of 1320 x 1035 spatial resolution and a pixel resolution of 12 bits. Cooling the camera reduces image noise and extends the system's dynamic range. Images can be captured and viewed on the Macintosh using IPLab software, Digital Image Processing Station, and IMAGE, a program written by Wayne Rasband of NIMH.

DCRT was originally encouraged in the report by the NIH Advisory Committee on Computer Usage to evaluate this technology and make it available at NIH. The system's high resolution makes it a superior tool for quantitatively evaluating biomedical images. CSL is assisting researchers in using the camera-computer system for imaging a variety of medical and biological mediums, primarily gels and film. Work in progress includes cataloging dental film and acquiring and processing 2-D gels with high resolution.

A variety of accessories are available for use with this camera including filters, lenses, lighting systems and mass storage devices. Researchers interested in digitizing gels can calibrate their images into density values with neutral density step wedges. Wavelength-selective Kodak wratten filters can be used with the system to enhance image contrast. The available lenses on the system can accommodate high resolution on the image plane down to 8" x 10" at lower spatial resolution. Reflective and/or transmissive light can be used.

During the coming year CSL anticipates using the camera for electrophoretogram digitization, film digitization, and microscopy. The best software for each technique will be examined and made available.

Acquisition and Analysis of Ophthalmic Images

M.A. Vivino, B.L. Trus, Ph.D.
with M. Datiles, M.D., S. Chintalagiri, B. Magno, M.D., S. Lassa, M.D., (NEI/OGCS)

The Clinical Branch (CB), NEI uses images produced by the Scheimpflug principle to quantitate eye opacities. The principal goal is to find a technique to accurately determine changes in a cataract lens and to eventually observe the effects of anticataract drugs. CB/NEI employs Scheimpflug Slitlamp Cameras (SLC) to visualize pathological changes in anterior

ocular measurements. A SLC measuring system consists of a special photo slit lamp with a video or photographic imaging system. In the Zeiss SLC, video data are acquired directly by a computer-based image analysis system. The relevant clinical information contained in the images is cataract area and density.

A year ago, CSL responded to a requirement to replace the data system for the Zeiss SLC. This data system was developed using a Macintosh II for image acquisition and processing, and an optical disc for image storage and retrieval. The existing Bosch Newvicon Video Camera was connected to Data Translation's Quick Capture Image Acquisition System. Quick Capture uses an 8-bit video ADC operating at 25 frames/second writing data to a 768 x 512 image buffer.

IMAGE, a program for general-purpose image analysis on Macintosh computers, developed by Wayne Rasband of NIMH, is used for acquisition and processing. Routines have been added to the IMAGE program for synchronous acquisition of images. The major calculations associated with eye image analysis are density measurements within regions. Regions are found using an automatic detection algorithm. Additional analysis is done with intensity profiles and histograms.

In addition to the Macintosh IMAGE program, software was developed on a VAX computer system to assist in the detection, quantitation, and classification of cataracts. The software may be useful in longitudinal studies involving anticataract drugs, and has been successful in showing the differences between different subclasses of cataracts. In pursuit of better methods for observing the cataract, we have performed a number of studies. These have included followup studies on the Zeiss and Topcon SLCs, reproducibility studies on the instrumentation used, and careful observation of clinical trials and statistical data from cataract studies performed around the world.

It is hoped that continued instrumentation development, image processing techniques, and CSL's close work with NEI physicians will eventually lead to clinical trials of pharmaceuticals that may prevent or reverse the cataract formation process. NEI, with CSL assistance, continues to examine new SLCs as they are developed and become available. Although the SLC imaging systems used in the NEI Clinic have been modified to give the clinically useful information NEI desires, a commercial system which performs these operations as an integral part of the system would be purchased if it were available. Some of the basic requirements for a new SLC would include the ability to produce images which can be calibrated to optical densities, a wide dynamic range, operation at low light level, a highly linear video system, distortion- and achromatic-free optics, and the ability to save images in formats adaptable to NEI analysis programs.

During the coming year, CSL will continue supporting NEI's cataract studies. Scheimpflug data from previous years will be analyzed using automatic region detection algorithms and the calibration of images to density values. Additional support for video signal acquisition and processing will be offered for corneal endothelium studies using a Keeler Specular Microscope, for retroillumination imaging using an Oxford and Neitz camera system, and for 2-D gel analysis of lens proteins.

Flow Cytometry Advanced Data Analysis (FC/ADA)

L. Barden
with R. Tate, Ph.D. (DCRT/CSL); S. Sharrow (NCI/DCBD/EIB); D. Plugge, J. Robinson (Systex, Inc.)

The Flow Cytometry Advanced Data Analysis Project (FC/ADA) is a laboratory automation project done in

collaboration with the Experimental Immunology Branch (EIB), Division of Cancer Biology, Diagnosis and Centers, NCI to design and implement a production-oriented research-class flow cytometry service facility capable of the acquisition, archiving and in-depth analysis of multiparameter flow cytometry data. A number of complementary analytical techniques, such as nonhierarchical cluster analysis and multidimensional histogramming, are applied to standard format data files. Experimental conditions and sample parameters are stored in machine-readable form along with the data.

A data staging and archiving system scaled to match the expected long-term data acquisition rates and storage requirements is being provided as part of the system. A relational database system will allow generalized searches of the stored data based on sample characteristics and experimental conditions.

When completed, this system will replace an existing data acquisition and analysis system currently in use in EIB that was also developed by CSL. While the direct collaborative effort involves CSL and EIB, the software and techniques developed under this project are shared with other flow cytometry facilities within the NIH intramural research program. Flow cytometry sites at NIAID and the FDA Center for Biologics Research are actively involved in this effort.

Current Status

The data acquisition system, a Becton-Dickenson FACSstar Plus flow cytometer with a VAXStation 3100-based data acquisition system and vendor-supplied software has passed acceptance testing in the EIB facility.

A data archiving system involving online disk storage, an autoloading magnetic tape carousel and archive management software is in place and has performed adequately under a test workload.

The cluster analysis program (CAP) is already in routine use in the EIB facility. In addition to performing nonhierarchical cluster analysis on list mode data files, this program is used to generate gated histograms from the same data.

The Laboratory Analysis Program (LAP), written in the C++ object-oriented programming language for use on Unix workstations as a general-purpose histogram and spectrogram analysis tool, has been successfully rehosted to run under the VAX VMS operating system as well.

A number of investigators at the EIB, NIAID and FDA sites are using PC- and Macintosh-based data analysis programs with data acquired at those facilities. This is made possible by the use of standardized formats for raw data and by the network connectivity available between the host facility and the investigator's offices and laboratories.

Plans for FY92

We expect to proceed with production use of new software and hardware at the EIB site early in FY92. The configuration should perform adequately under a production workload, but we are not yet finished with the work necessary to provide EIB with operational independence from CSL. In addition, the preproduction use of existing programs has pointed up areas in which additional effort will pay significant dividends, both for the EIB site and for other cooperating sites in the NIH intramural research program.

Once the EIB facility is able to function on a production basis with the VAX/VMS-based data acquisition, archiving and analysis systems, we will begin implementing an online relational database containing selected textual information from the standard

flow cytometry data files. Fully populating the database would allow contextual searches over the entire archive of stored data.

The cluster analysis program, CAP, has been in general use for almost a year now. Although the current version performs nonhierarchical cluster analysis adequately, we have found several areas that are appropriate for further enhancement. We will proceed with further development of this program, concentrating on making significant use of the network-based computational services available from the Convex minisupercomputer and the NIH Highly Parallel Computer System and on expanding the algorithmic flexibility of the program.

The histogram analysis program, HAP, provides a flexible implementation of generalized spectrogram and histogram data analysis for Unix and VAX/VMS computer platforms. Our efforts in this area will concentrate on providing adequate user documentation, optimizing for speed of execution in a VAX/VMS timesharing environment, and enhancing ease of use of the program.

As resources permit, we will work to encourage an existing trend among our more sophisticated users to perform specialized analysis of their research data, acquired and archived at a primary site, on networked PC and Macintosh computers located in their own offices and laboratories.

Laboratory Application Package (LAP)

J. Powell

The Laboratory Application Package (LAP) was originally developed to run on Unix workstations as a general-purpose tool for the analysis of spectrophotometric data. LAP is already extensively used by approximately 40 researchers located primarily in two NIDDK laboratories in Building 2. During FY91, the program (written in the C++ object-oriented programming language) was ported to the VAX/VMS operating system and was partially tailored for the analysis of flow cytometry data.

LAP can perform a wide range of data manipulations on vector data, xy paired data, and matrix data using either a command or expression syntax. Customized command procedures can be saved in files and added to the LAP/command set. Results may be viewed as line graphs, scatter graphs, bar graphs, perspective views, or contours in color or monochrome viewports within an X window or on Tektronix 4010 and 4107 compatible terminals. Hardcopy graphical output (both monochrome and color) is supported on HPGL and PostScript plotters.

During the past year LAP was ported successfully to the VAX/VMS environment for use by the flow cytometry community. However, it needs further improvements in terms of functionality and performance to meet the requirements of production flow cytometry laboratories. We will work toward completing these improvements in the coming year. In addition, LAP will be thoroughly documented with a general reference manual covering LAP commands and with individual users manuals for flow cytometry and spectrophotometric applications. A port of LAP to the Convex will proceed as well. After this Convex LAP has been adequately tested and documented and a support mechanism has been established, it will be made available to users on the Convex.

Laboratory Data Acquisition

M.A. Vivino
with S. Yoshikami, Ph.D., W. Hagins, M.D., Ph.D. (NIDDK/LCP)

The Laboratory of Chemical Physics (LCP), Membrane Biophysics Section, NIDDK, uses computer-controlled

instrumentation developed by CSL in their research on phototransduction mechanisms in visual cells. The host Macintosh computer selected for this system is capable of data acquisition with simultaneous instrument control, and display and preliminary analysis of data, and can operate as a terminal for analysis and visualization on networked computer systems.

National Instruments LabView II was chosen as the data acquisition and control software for the Macintosh. LabView implements graphical programming; a diagram on the screen represents the source code for the instrument. National Instruments describes this type of design as a "Virtual Instrument" (VI). The VI consists of a front panel and block diagram displayed on the Macintosh. The user interacts with the front panel using computer-generated controls, digital displays and graphs during instrument operations. The block diagram is analogous to source code (a computer program). It has looping structures and logical operations that can be varied by interacting with the display using the standard Macintosh features. LabView interprets the block diagram "source code" for sequencing VI operations. LabView's flexibility and the ease of user modifications should prove useful for experimental research labs.

For this system, LabView II is configured for controlling, digitizing, and saving experimental data, with accompanying laboratory notes, on experimental conditions and observations. The signals acquired are electrical and temperature signals issued simultaneously from temperature and voltage transducers. The transducers are located in an instrument designed in LCP for microcalorimetery and electroretinography. Isolated vertebrate retinas are maintained under physiological conditions within the instrument. LabView II controls light flashes of varying intensities and frequencies to initiate phototransduction in the visual cells. The evoked microdegree temperature changes arising from light-simulated biochemical and electrical activities in the tissue, and the light-triggered electrical currents, are captured into the LabView II data acquisition system.

CSL will continue to collaborate with the Membrane Biophysics Section in applying the Macintosh/LabView II to capture and process data from experiments in phototransduction mechanisms.

Computer-Assisted Patient Interviewing in Clinical Pharmacy

J.M. DeLeo
with F. Pucino and K. Calis (CC/Phar. Dept.)

Clinical pharmacists are being increasingly called upon to participate more directly in patient care. They continue to become more deeply involved in providing pharmacotherapeutic information to patients, assessing patient compliance with their medication regimens, screening for adverse clinical events linked to single drugs and various drug combinations, and sorting out possible complications between medications and food, allergies, and medical conditions. To be fully effective in these demanding tasks, pharmacists must keep abreast of new drug information, allocate more one-on-one time for patients, and continually sharpen their interviewing skills.

To explore potential roles for the computer in assisting with these emerging tasks and skill requirements, a collaborative project began between CSL and the Clinical Center Pharmacy Department in January 1990 to design and develop a computer system that would directly interview patients. The objective is to develop a computer interviewing system to collect medication histories and to flag possible untoward effects related to medication regimens, thereby making more pharmacist time available for patients

who are not candidates for direct computer interviewing. Additional design objectives included flexibility in authoring interview scripts, maintaining online updated comprehensive information on drugs and adverse clinical events, and making the computer interviewing available in the NIH Pharmacy outpatient waiting area as well as in a variety of clinical settings supported by the Pharmacy Department.

The initial version of the patient interviewing system was developed and tested in late FY90. This version performed the medication historytaking task. The interview scripts for that system were designed to solicit the following information: patient description, medical conditions, medication regimen, medication compliance, symptoms and implied medications, allergies, dietary history, social history, occupational and environmental toxic exposure, and patient evaluation of the system. A concise report was generated and produced at the conclusion of the interview for the attending physician. The knowledge gained with this experience suggested major enhancements in script content and style.

The United States Pharmacopeial Convention Drug Information database was chosen for online use with the patient interviewing system. It was chosen because it contains adverse clinical events in lay terminology. Using this feature, we have built a lay-terminology thesaurus that provides natural-language interfacing between patient input and the drug information database. Also, a comprehensive lay-terminology adverse-events symptoms survey allows questioning of patients on possible adverse effects of medications. A spinoff of the thesaurus development work will be used by USP to standardize their adverse-effects terminology and eliminate redundant terms for describing the same adverse event. We have also developed a machine-learned ADR detection and evaluation scheme using Delphi polling results obtained from 10 clinical pharmacists. This is being integrated into the system.

During the coming year, we will refine the thesaurus, develop a method for extracting data from the USP natural-language database, complete system integration, and begin preliminary testing of the integrated system.

Breast Cancer Patient Survival Prediction—A Neural Network Approach

J.M. DeLeo
with G.R. Merlo (NCI/DCBD/LTIB)

Breast cancer survival prediction is presently based on several traditional prognostic factors such as tumor size, lymph node involvment, estrogen and progesterone receptor status and histologic grade. New prognostic factors such as proliferative index, cathepsin D, genetic mutations, and expression of tumor-associated antigens are under study in predicting outcome for node-negative patients. Paradoxically, as new factors prove useful, incorporating these factors with traditional methods in clinical decisionmaking becomes increasingly difficult because of the parameter combinatoric complexities. Neural networks have been highly successful in analogous multidimensional pattern classification problems in many engineering applications. These new biologically inspired computer architectures that adaptively learn to classify complex patterns of data may be useful in breast cancer relapse and survival prediction and related clinical decisionmaking. Furthermore, new proven prognostic factors should be easily incorporated within neural network methodology.

We developed a back-error propagation neural network to predict patient disease-free survival using the classical, well-established parameters listed above

to explore these ideas. Training and testing the network were performed with clinical data derived from 170 cases of breast cancer. We are grateful to Drs. Cappa, Liscia and Gaglia of the S. Giovanni Hospital in Torino for collecting and sharing their data. Software to model the back-error propagation paradigm was written by the principal investigator with a PC-compiled BASIC platform. Several concerns arose when transporting neural network technology to clinical medicine classification problems such as breast cancer relapse and survival prediction. These concerns include: sparseness of data usually encountered in medical applications and the need for confidence measures and explanations in medical decisions—information neural networks do not usually provide. Once these concerns are properly addressed, neural networks may provide powerful new tools to assist the oncologist in selecting the optimum therapy plan for the individual patient, especially the node-negative patient for whom relapse prediction is critically important.

Our experience has led us to several conclusions:

- input parameters should be scaled
- weights should be computed at each significant time point
- improved weight convergence methods are needed
- single patient survival vs. time plots are desirable outputs
- confidence intervals are needed and can be computed with bootstrapping methods
- larger databases are needed to establish meaningful confidence intervals
- higher performance computing is needed to process larger data sets and confidence intervals
- postcensor survival probabilities in training data can be completed using Kaplan-Meier survival analysis methods.

These findings were presented at the Biennial Meeting of the International Association for Breast Cancer Research. Several offers for collaboration were made by various university and cancer center representatives, most of whom had large databases to share.

The Lipid Analysis Sample Tracking System (LASTS)

R.L. Tate, Ph.D.
with J. Hoeg, M.D., D. Wood (NHLBI/MDB)

Over the past 20 years the Molecular Disease Branch (MDB) of the NHLBI has studied lipid metabolism disorders by analyzing tens of thousands of blood samples from nearly 7,000 individuals. Until recently, all data was gathered and entered by hand into central NIH computer utility databases. As the volume of samples and corresponding staff workloads rose, it became increasingly difficult to process and enter data in a timely way. Furthermore, automated blood chemistry analyzers and powerful personal computer database software pointed the way to a comprehensive automated method for acquiring and processing the analytical results and tracking the samples.

LASTS is a comprehensive PC-based system for recording the results of lipid analyses performed on plasma samples submitted to the MDB's Lipid Analysis Laboratory. Identifying information about the samples is entered into databases maintained on a laboratory PC and verified when the sample is acquired. The samples, identified by bar-coded labels, are subdivided for analysis. The system maintains records of the number of samples awaiting each type of analysis, scheduling appropriate test runs when a sufficient number of samples have accumulated.

Analyses are performed in microplates manipulated by a robotic sample handling system that includes a microplate colorimeter interfaced to the LASTS sys-

tem. As each analysis is performed, the results are either captured directly from the analyzer or keyed by the bar-coded label for manual entry. The results to date on each sample are maintained in a database that can be searched by laboratory personnel or the referring physician. Once the validity of the test results have been certified, the sample data are copied to a report dataset that is then transferred to the NIH Central Computer Utility and incorporated into the MDB lipid study databases. Verified data is also maintained locally in a form suitable for access by PC-based database query programs. Statistics about controls and standards are also maintained.

The full LASTS system is now in test use in parallel with the manual methods. Early indications are that LASTS permits a three- to five-fold increase in sample throughput and a drastic decrease in manual recordkeeping. Full replacement of the manual methods is expected to take place in the first quarter of FY92.

Neuromagnetometer Computer System

R.L. Martino, Ph.D.
with C.A. Johnson, T.K. Yap, S.T. Phan (DCRT/CSL); S. Sato, M.D., M. Balish, M.D. (NINDS/Office of the Clinical Director)

The Medical Neurology Branch, NINDS, and the Computer Systems Laboratory, DCRT, have collaborated on a research project to noninvasively localize epileptic discharge sources within the human brain by using neuromagnetic recording in conjunction with conventional electroencephalogram (EEG) recording. Many patients with seizure disorders exhibit low-level cellular discharges between seizures, indicated by interictal spikes or sharp waves in their EEG and magnetoencephalogram (MEG) recordings. This project involved the development of computer techniques for automating and enhancing the procedure that is used by NINDS neurologists to determine the intracranial locations of the sources of epileptiform discharges in patients with epilepsy.

In a previous fiscal year, CSL designed a computer system that detects epileptiform discharges from the EEG and MEG signals in realtime. This system has been implemented with a number of detection algorithms, including both those developed by CSL and some selected from the published literature, and a variety of options for configuring these algorithms, allowing the medical staff to choose the optimal method for a given patient. The system provides a real-time display of the signals showing where an event is detected, and allows the neurologists to manually save both epileptiform discharges and longer seizure activity.

During the past year, CSL performed a comprehensive clinical evaluation of the real-time detection system with a variety of patient signals, and compared its performance to that of a new commercial system that uses an algorithm not available on the CSL system. The CSL system reliably detected epileptic events, performed better than the commercial system, and was placed into clinical operation.

Diode Array Spectrophotometer

H.A. Fredrickson
with A.R. Schultz (DCRT/CSL); W. Frauf, J. Cole, P. Smith, Ph.D. (NCRR/BEIP); R.W. Hendler, Ph.D. (NHLBI/IR/L)

A computer-controlled 100-Channel High-Speed Spectrophotometer is being developed by the Biomedical Engineering and Instrumentation Program of the National Center for Research Resources and CSL for the Laboratory of Cell Biology, NHLBI. It will be used to obtain more complete spectral infor-

mation concerning the rapid changes of the reduction and oxidation centers within the protein enzyme cytochrome oxidase. This enzyme is involved in cellular respiration and is located within the inner lipid bilayer of the mitochondria.

The electronic hardware will consist of two 50-element photodiode arrays connected to 100 discrete A/D converter and local storage channels. Each channel is capable of acquiring data every 10 microseconds. A personal computer (PC) will be used to control the spectrophotometer. Timing control sequences will be transmitted from the PC, and the PC will receive data from the 100 A/D channels via a 40-bit parallel interface.

CSL has assisted in defining the interface between the spectrophotometer and the computer, and has developed the data acquisition and control software. The researcher intends to use the modeling program MLAB to analyze the data.

BEIP has designed, built, and, with the assistance of CSL, tested a prototype eight-channel interface board with one analog-to-digital converter and its associated 1,024 words of high-speed memory per input. Subsequently, they arranged for this design to be fabricated in a printed circuit board. A "mother board" capable of accepting and controlling 12 interface boards has also been fabricated as a printed circuit. The spectrophotometer is in a similar stage of fabrication. CSL is working with BEIP on a daily basis to assemble and test the complete interface. This project should be completed early in FY92.

VAX/VMS Services

R.L. Tate, Ph.D.
with L.K. Barden (DCRT/CSL)

The number of Digital Equipment Corporation VAX computer systems using the VMS operating system on the NIH campus has grown to the point that there are now nearly 100 VAX/VMS systems at NIH that provide computing services to over 800 users. In recognition of the importance of these systems to the NIH scientific community, CSL developed a service to enhance VAX and VMS support, consolidating support previously supplied to other projects and extending support to new and continuing activities throughout NIH.

An outgrowth of support for VAX/VMS-based projects within CSL, this activity provides hardware and software support resources for VAX/VMS computing operations throughout the NIH. Hardware, software tools, such as language compilers, database systems, connectivity products; and consultation services form the basis for this support. The hardware support is provided by a two-member VAX cluster with an aggregate of 48 Mbytes of memory, 3 Gbytes of disk space, several streaming tape and CD-ROM drives, a fast-floating point accelerator subsystem, and serial line and Ethernet communications controllers. Software support covers FORTRAN, Pascal, C, and C++ compilers, several code development support utilities, network diagnostic utilities, distributed file and queuing service software, Rdb database development tools, screen and hardcopy graphics support libraries, and a large library of user-contributed software.

Consultative services are furnished primarily through the "VMS Hotline," a contractor-operated adjunct to consultation provided by the CSL staff. The Hotline, which is staffed during regular business hours, permits intramural VMS users and systems personnel to make telephone requests for assistance with common system operations, such as software installation or configuration changes, as well as to obtain help with problems. The requested services are furnished by telephone, by electronic mail, or at the customer's site, whichever is most appropriate.

An important part of this activity is sponsorship of the Biomedical VAX/VMS Users' Group, devoted to the discussion of topics of specific interest to VAX/VMS users and system managers in the biomedical community. The membership of this group, which meets bimonthly, represents almost all the major VAX/VMS sites on the NIH reservation as well as several in off-campus locations.

DECnet Management

R.L. Tate, Ph.D.

DECnet usage of the NIH campus-wide network continued its expansion this past year. Over 175 systems using the DECnet networking protocol are now connected. Coordination of this expansion among the various laboratories and IDCs involved resulted in virtually troublefree growth. The principal investigator for this project provides the central coordination necessary for smooth incorporation of new DECnet hosts into an integrated network that spans both the NUnet and the RESnet portions of the campus network. Because of the increasing popularity of DECnet among PC and Apple Macintosh users on LANs served by VAX/VMS systems, the estimated number of interconnected DECnet nodes will grow to over 300 in the coming fiscal year.

PHS Telecommunications Project

A.R. Schultz

The DHHS Public Health Service is in the process of procuring a state-of-the-art digital telephone system. NIH had four representives on the Technical Evaluation Committee, including two from DCRT. The procurement is for a digital switch/telecommunications management system to replace Centrex service in 18 locations serving 22,000 lines. NIH has approximately 7,700 voice lines and 5,200 data lines. The technical evaluation began in March 1989 with 4-hour meetings scheduled 3 days per week. The evaluation was completed including several site visits during the fall of 1990. DCRT proposed and wrote several sections of the contract that would enable the installation of telecommunication media for implementing higher speed networks at NIH. When the contract is awarded, it will include engineering, installation, testing, placing in service, training of government personnel, and ongoing operation and maintenance of a complete "state-of-the-art" integrated telecommunication system.

Publications and Presentations

Aldroubi A, Unser M, Tietz D, Trus B. Computerized methods for analyzing two-dimensional agarose gel electropherograms, *Electrophoresis* 1991;12:39-46.

Booy FP, Newcomb WW, Trus BL, Brown JC, Baker TS, Steven AC. Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus, *Cell* 1991;64 1007-1015.

Corso DM, DeLeo JM, Pucino F, Calis KA, Gallelli JF. Development of a comprehensive symptom questionnaire for detecting potential adverse effects, Tenth Annual Eastern States Conference for Pharmacy Residents and Preceptors, Baltimore, April, 1991 (presentation).

DeLeo JM, Pucino F, Calis K. Portable patient-entry medication history computer system, American Association of Hospital Pharmacists, San Diego, June,1991 (poster).

DeLeo JM, Pucino F, Calis K. Computer-learned adverse drug reaction detection, American Association of Hospital Pharmacists June, 1991 (presentation).

Deleo JM, Campbell G. The fuzzy receiver operating characteristic function and medical decisions with uncertainty, First International Symposium on Uncertainty Modeling and Analysis, College Park, Maryland, December, 1990.

DeLeo JM, Merlo GR, Gaglia P, Liscia, DS. A neural network approach to survival prediction in breast cancer patients, Biennial Meeting of the International Association for Breast Cancer Research, Aosta Valley, Italy, May, 1991.

Edwards PA, Datiles MB, Unser M, Trus BL, Freidlin V, Kashima K. Computerized cataract detection and classification, *Curr Eye Res* 1990;6:517-524.

Kempner KM, Ostrow HG, Fessler JF, Tucker EE. Prototype system for timely image distribution in support of a cardiology outpatient clinic. In: Murray A, ed. Proceedings of the 1991 Computers in Cardiology Conference, Institute of Electrical and Electronics Engineers, Inc. New York (in press).

Kocsis E, Trus BL, Steer CJ, Bisher ME, Steven AC. Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment, *J Struct Biol* 1991 (in press).

Kocsis E, Trus BL, Steer CJ, Bisher ME, Steven AC. The clathrin triskelion leg has an extensible proximal domain. In: Peachey LD, Williams DB, eds. *Proceedings of the XIIth International Congress for Electron Microscopy*. San Francisco, 1990;3:468-469.

Martino RL, Johnson CA, Yap TK, Sub EB, Pfeifer JC, Hodoscek M, Brooks BR. Highly parallel computing at the National Institutes of Health. In: Parker GV, ed. *Proceedings of the Supercomputing USA/Pacific 91 Conference*. Santa Clara, 1991;110-111.

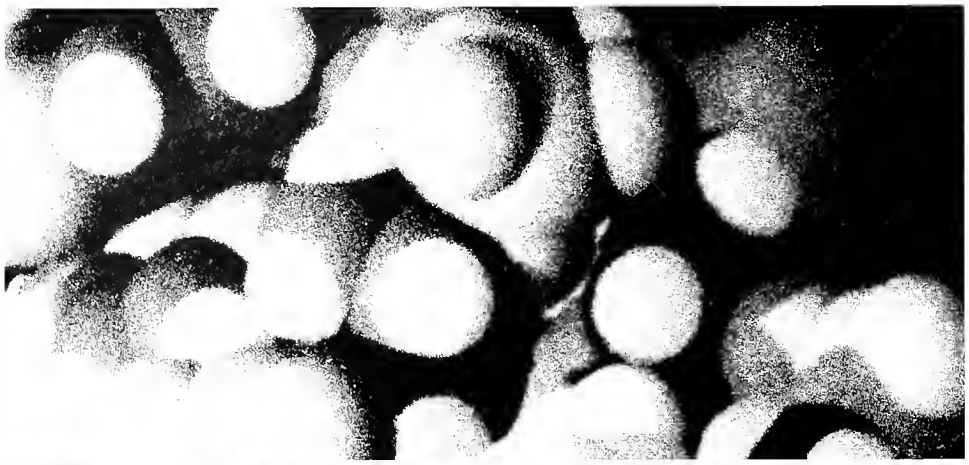
Palmeri ST, Kempner, KM, Power JA, Bacharach SL, Choi BW, Rosing DR, Bonow RO. Effects of percutaneous transluminal coronary angioplasty on exercise-induced changes in R-wave amplitude, *Am J Cardiol* 1991;68(1):114-116.

Steven AC, Kocsis E, Unser M, Trus BL. Spatial disorders and computational cures, *Int J Biol Macromol* 1991 (in press).

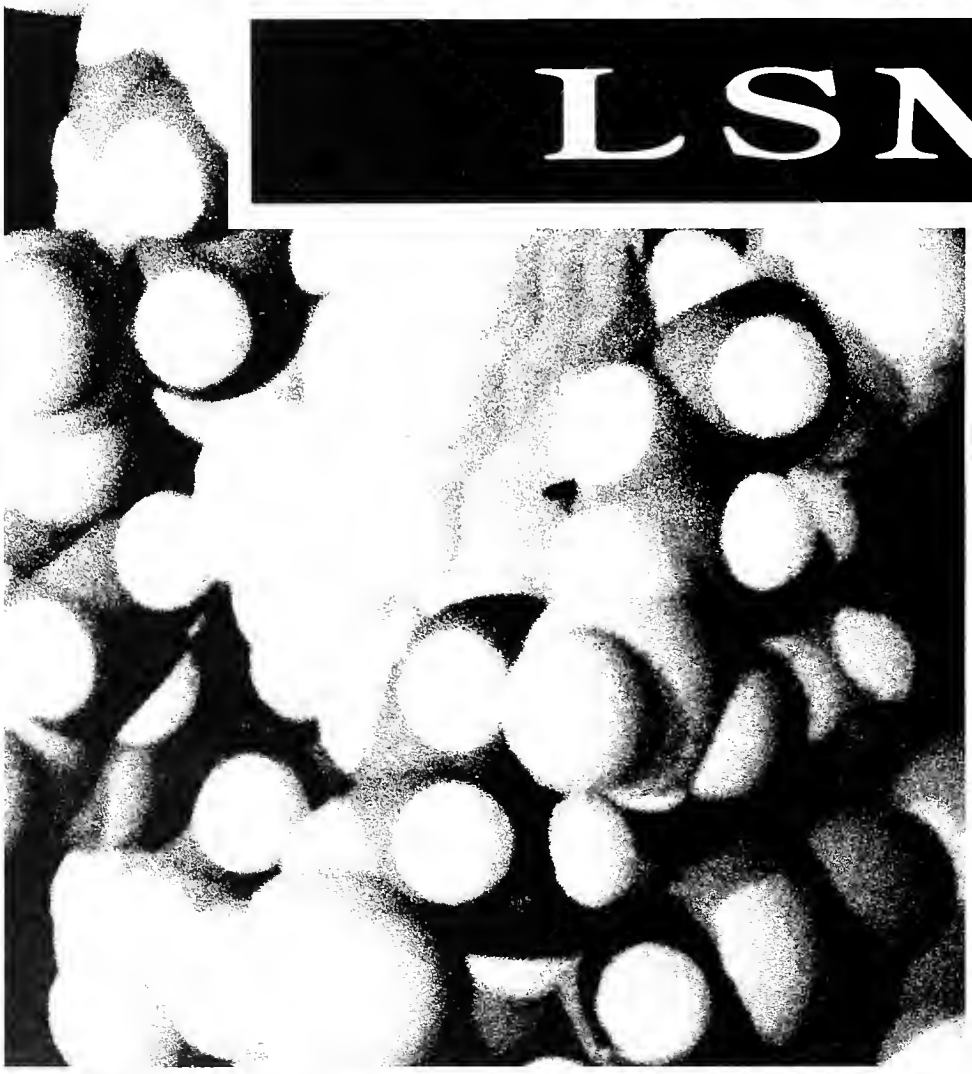
Steven AC, Mack JW, Trus BL, Bisher ME, Steinert PM. Structure and assembly of intermediate filaments: multi-faceted, myosin-like (but non-motile) cytoskeletal polymers. *Cytoskeletal and Extracellular Proteins*, 1991;15-26.

Thomas DJ, Wall JS, Hainfeld JF, Kaczorek M, Booy FP, Trus BL, Eiserling FA, Steven AC. gp160, the envelope glycoprotein of human immunodeficiency virus type 1, is a dimer of 125-kilodalton subunits stabilized through interactions between their gp41 domains. *J Virol* 1991; 65:3797-3803.

Thomas L, Kocsis E, Colombini M, Erbe E, Trus BL, Steven AC. Surface topography and molecular stoichiometry of the mitochondrial channel, VDAC, in crystalline arrays. *J Struct Biol* 1991;106:161-171.



LSM



Laboratory of Statistical and Mathematical Methodology

Gregory Campbell, Ph.D., Acting Chief

Executive Summary

The interface of medicine, statistics and computers continues to drive and focalize the Laboratory of Statistical and Mathematical Methodology's (LSM's) mission. This year's highlighted efforts demonstrate the variety of ways in research, consulting and service that LSM contributes to progress at NIH and to biomedicine in general.

An LSM research area of importance in medicine concerns Receiver Operating Characteristic (ROC) curves to evaluate diagnostic tests. Great strides in the statistical theory associated with ROC curves have been made by LSM. New statistical and computer-intensive procedures using methodologies such as the bootstrap have been developed at LSM to compare two or more tests in terms of sensitivity and specificity performance. The laboratory has applied these powerful procedures to a number of collaborative studies, including the evaluation of left ventricular hypertrophy using the Framingham Heart Study data and the evaluation of Cushing's syndrome.

A second area of LSM research that is finding immediate applications in biomedicine focuses on statistical procedures for missing and censored data using the Expectation-Maximization (EM) algorithm. LSM research in and applications of this methodology (important in statistical analyses associated with variance components and in Gibbs sampling) is currently under way. An associated LSM effort is a study of the statistical implications of artificial neural networks and the performance of such compared to more classical statistical prediction methodologies. These highly computer-intensive procedures lend themselves well to parallel machine implementation; a program to do this has been launched by LSM.

In the consultation and collaboration arena, LSM has been particularly active. One example is the noteworthy collaboration with a National Cancer Institute laboratory conducting a molecular biological study of the DNA mutations in the tumor tissue of breast cancer. New statistical techniques were pioneered by LSM with this group to identify the correlation of loss of heterozygosity in the regions on different chromosomes.

Concerning service to the NIH, this year records the start of several exciting efforts. LSM is moving swiftly on a number of fronts, keeping pace with the dramatic improvements in mathematical and statistical software. A seminar was offered by LSM staff to over 175 people concerning the transition to the latest, most powerful version of SAS, the mainstay of mainframe data management, statistical analysis and computer-generated reports at the NIH. LSM-negotiated site licenses for statistical products on DOS-based machines have expanded. Compendia of PC and Macintosh statistical software reviews have been generated by LSM and made available to the biomedical community. LSM support has begun for Mathematica, a multiplatformed product for effortlessly doing mathematical and symbolic computation by computer. A new, LSM-spearheaded course, Macintosh Software for the Scientist, has been well received and has helped establish LSM as the primary source of expertise for statistical, mathematical, graphical, and scientific typesetting programs on the Macintosh platform.

The Laboratory of Statistical and Mathematical Methodology (LSM) combines research in mathematical statistics, mathematics, and computer and information science with collaboration and service in these areas for NIH researchers and administrators. LSM staff interact with all NIH Institutes, other Federal agencies outside DHHS, and biomedical researchers worldwide.

LSM has 13 fulltime professionals, including research mathematicians, mathematical statisticians, computer systems analysts, and programmers. There are three sections:

- *Statistical Software Section (SSS)* provides consultation to and collaboration with NIH research-

ers and administrators in all computational aspects of biomedical data analysis, including selection and support of large systems and packages.

- *Statistical Methodology Section (SMS)* conducts independent research in mathematical statistics with applications to biomedicine, provides expert statistical consultation and collaboration with biomedical researchers and administrators, develops software tools for DNA analysis and provides support to the Statistical Software Section in the statistical aspects of computer software and their evaluation.
- *Biomathematics and Computer Science Section (BCS)* performs independent research, provides consultation in mathematics and on software used for scientific printing, and develops LSM computer network facilities.
- A fourth section, the *Medical Information Science Section (MIS)*, investigated and developed methods for applying information and computer science to medical language data processing. While this section has been inactivated in May 1991, a major computer program developed by it is still in routine use by the Clinical Center, NIH.

Computation in FY91

LSM activity offers biostatistical, biomathematical, and other scientific systems and packages to the NIH user community and evaluates new systems and packages for suitability to NIH needs. Computer systems and packages supported by LSM are shown in Table 1. Use of mainframe statistical packages at NIH since 1975 is shown in Figure 1, on page 64. Average monthly use remained at a high level in FY91.

As in previous years, the SAS statistical and data management system was extensively used at NIH, with an average of 85,800 accesses per month via the System 370. SPSS was accessed around 3,500 times per month, and the BMDP package was accessed an average of 425 times per month.

LSM mainframe statistical support included maintenance of the system or package and provision of adequate documentation, including NIH computer system changes, system or package updates, and corrections. It also included rapid response to queries concerning user access to the most used systems and packages. The SSS staff answered over 3,500 calls for SAS assistance, including requests for information on job control language, program parameters, and other operating system procedures, as well as assistance in interpreting results.

Other mainframe software supported by LSM had more limited use. Usage ranged from an estimated 1,000 MLAB sessions, 400 GRAPH sessions (until the discontinuation of the DECsystem-10 service in February 1991), and 200 VMAP accesses per month to relatively few sessions for such specialized programs as GLIM and CART. Support for IMSL has expanded to include the Convex as well as System 370 mainframes.

While statistical activity on MS-DOS and Macintosh microcomputers is more difficult to assess, LSM has continued to expand its support of software on such rapidly growing platforms. The usage of LSM-negotiated NIH site licenses for Base SAS, SAS/STAT and SAS/GRAPH has continued to expand and two other products, SAS/IML and SAS/FSP, have been added. A Blanket Purchase Agreement for all of the NIH was negotiated for the product BMDP-PC. Support for these products and for SPSS/PC and IMSL has continued, and support for several Macintosh packages has begun. On-going evaluation of other statistical packages on MS-DOS

Table 1. Systems and Packages Supported by LSM

SAS, SAS/GRAPH, SAS/ETS, SAS/OR, SAS/FSP, SAS/AF, SAS/IML, SAS/CBT
Vendor: SAS Institute, Inc. A batch and interactive System 370 system for statistical analysis, with extensive file manipulation capabilities and graphics, also in interactive mode on MS-DOS machines.

SPSS, SPSS-PC
Vendor: SPSS, Inc. A system for univariate and multivariate statistical analysis with file handling capabilities, in batch mode on the System 370, and interactive mode on System 370 and MS-DOS machines.

BMDP
Vendor: BMDP Statistical Software, Inc. A collection of System 370 batch programs for univariate and multivariate statistical analysis.

IMSL (International Mathematical and Statistical Libraries)
Vendor: IMSL, Inc. An extensive collection of FORTRAN routines for statistical and mathematical analysis for System 370, Convex and MS-DOS machines.

Mathematica
Vendor: Wolfram Research Inc. A computing environment for doing mathematics with numerical and symbolic calculations and graphics, available on Macintosh, Convex and MS-DOS machines.

MSTAT 1
Source: DCRT staff. System 370 batch programs and subroutines for mathematical and statistical analysis.

GLIM (Generalized Linear Interactive Modeling)
Vendor: Numerical Algorithms Group, Inc. System 370 batch and interactive system for analysis of linear statistical models.

CART (Classification and Regression Trees)
Vendor: California Statistical Software, Inc. System 370 batch and interactive system for tree-structured regression and classification analysis.

DNAdraw
Source: LSM staff. A public-domain menu-driven MS-DOS program for preparing publishable displays of DNA sequences.

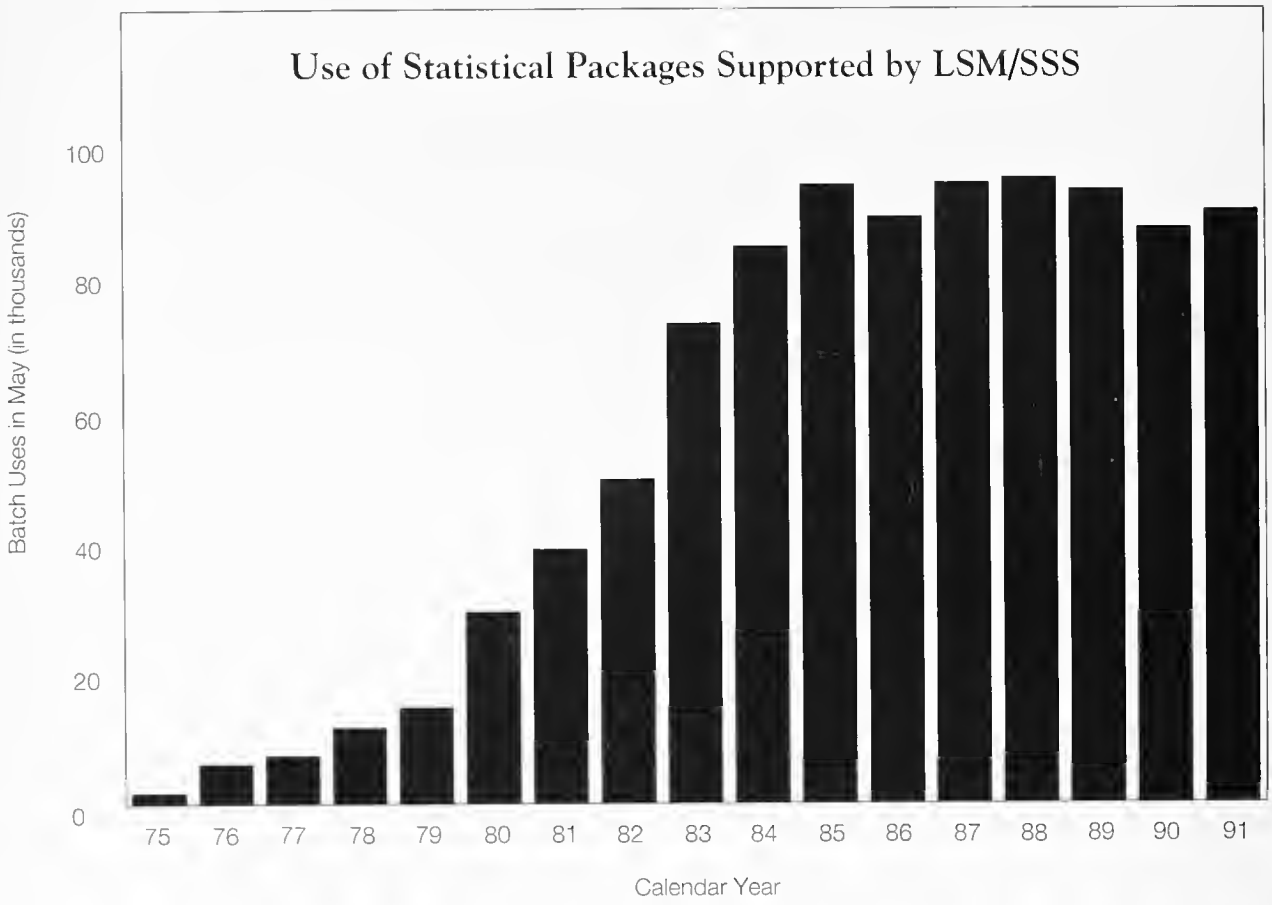
SPU (Scientific Printing Utility)
Source: LSM staff. A public-domain, MS-DOS program used for printing scientific text and diagrams. The related System 370 batch program, called VMAP, is used in conjunction with the WYLBUR text editor for IBM 6670 laser printers.

FPU (File Printing Utility)
Source: LSM Staff. A public-domain, MS-DOS program that offers easy formatting and fonts. It is especially useful for printing downloaded SAS output.

and Macintosh platforms has intensified. Collections of published review articles of statistical software have been compiled by LSM for use in the DCRT library. Support for the software product Mathematica has begun for MS-DOS and Macintosh microcomputers. Versions of the LSM-developed utilities SPU and FPU have been completed, documented and distributed. A new version of DNAdraw for IBM-PC compatible microcomputers that formats, annotates, and displays DNA sequences for publication was distributed in FY91.

Recognizing the importance of teaching effective use of systems and packages to biomedical researchers and other NIH users, LSM maintained a substantial program of short courses, documentation preparation, and informational talks and articles. SSS taught eight sections of the introductory course on SAS, three on SAS/GRAPH, one on Permanent SAS Files, and two

Figure 1.



sections of the introductory course on SPSS. SSS also presented two sections of a seminar on SAS 6.06 to 150 experienced SAS users. Seminars for DNAdraw were also offered, as well as three sections of the statistical course, "Recurrent Problems in Data Analysis." LSM helped to coordinate a new course on Macintosh software for the biomedical scientists. A new LSM course for the mathematical and symbolic computation software program, Mathematica, was introduced.

Consultation and Collaboration in FY91

LSM provided consultation in a wide range of scientific fields and in interpreting the results of statistical and other scientific computation. Staff members also provided consultation in biostatistics, biomathematics, and computer analysis of DNA sequences. Consultations were brief or extensive, depending on the complexity of the statistical or scientific research; some evolved into long-term collaborations resulting in scientific publications.

LSM's consultation services have been well utilized by many NIH components. For example, consultations on the use of the DNAdraw program for producing publication quality drawings of DNA sequences included work with 12 investigators from 9 NIH institutes and centers. Biomedical consultations arising from the mathematical software, Mathematica, were conducted with two other NIH investigators.

Collaborative statistical efforts included the following:

- R. Callahan, C. Cropp (NCI/LTI). For a study of loss of heterozygosity (LOH) using restriction fragment length polymorphisms for various regions of the human genome in a panel of tumor tissue of women with breast cancer, LSM provided basic analysis and a novel methodology to study LOH at various sites. In particular, the novel analysis was able to detect a low level of dependence among oncogene sites on different chromosomes, suggesting that certain combinations of mutations may collaborate in the development and malignant progression of breast carcinomas.
- K. Bayne (NCRR/OD). The behavior of caged rhesus monkeys (*Macaca mulatta*) was the subject of two studies where it was hypothesized that the introduction of environmental enrichment devices might reduce abonormal behaviors and foster grooming and foraging. LSM provided a novel nonparametric analysis to evaluate the effectiveness of placing pieces of fiberglass covered with artificial fleece and astroturf into the cages. The nonparametric analyses confirmed that these devices significantly reduced aberrant behavior and increased foraging/grooming activity levels.
- D. Rabin (NICHD/DEB). For research comparing normal women and women with Premenstrual Syndrome (PMS), LSM provided statistical advice, including stepdown multivariate procedures, for analyzing an experiment that studied the response of adrenocorticotropic hormone (ACTH) and cortisol to ovine corticotropin releasing hormone (oCRH) in both luteal and follicular phases. The resulting negative correlation between time integrated plasma ACTH and cortisol response to both oCRH and basal luteal progesterone concentration that was present in normal women but not in PMS patients suggested that women with PMS might have transient episodic disturbances in the hypothalamic-pituitary-adrenal axis.
- S. Marx (NIDDK/MD). In a study of patients with hyperparathyroid hyperplasia, an LSM analysis enabled the investigator to evaluate the size and shape of the four hyperparathyroid glands based on approximated dimensions, shape and weight. The purpose was to see if the sizes of the four glands were symmetric horizontally or vertically and to quantify gland-size variability. The study included estimation of volume on clay models by two observers to quantify reader variation. These were analyzed using log volumes and ratios of extreme values. The insight gained here will be applied in the surgical removal of glands of hyperthyroid patients.
- M. Flack, G. Cutler (NICHD/DEB). A standard test, using 17-hydroxysteroid excretion, was compared with urine-free cortisol from the dexamethasone suppression test in 188 patients with surgically confirmed Cushing's syndrome. LSM provided the statistical analysis involving Cochran's Q statistic and receiver operator curves (ROC) to evaluate these tests. A combination of the two tests was shown to decrease the likelihood of unnecessary transsphenoidal surgery while identifying a subpopulation of patients who required more invasive tests.
- S. Yanovski, J. Yanovski (NICHD/DEB). In an intensive study of the telephone (on-call) decision-making process for primary care family practice and pediatricians, three scenarios were presented to 142 physicians in a prearranged "on-call" setting. The responses (final correct diagnosis, number of critical areas addressed, and time to decision) were examined using contingency tables, parametric and nonparametric analyses of variance, and Goodman and Kruskal's gamma statistic as a measure of regression on ordinal category data. These statistical procedures revealed that medical faculty outperformed residents

and private practitioners in evaluating a case of severe dehydration. Overall, faculty members performed better in identifying severely ill children without misevaluating the less ill.

- K. Oerter and G. Cutler (NICHD/DEB). An improved diagnostic test for assessing hormone growth deficiency in children was devised in collaboration with LSM. Since mean spontaneous nighttime growth hormone (GH) levels were not shown to be dependent on IGF-1 levels, LSM constructed an alternative test that was a combination of the two. Application of this test to 51 normal children and 80 children with short stature demonstrated that it was significantly more sensitive than the spontaneous GH test and that, unlike the stimulation test, it was highly specific in the normal population.
- N. Martin, R. Rubinfeld (George Washington University). This study examined the link between giant papillary conjunctivitis (GPC) and meibomian gland dysfunction (MGD). To assess the need for vigorous treatment of MGD in patients with GPC by contact lens practitioners, LSM applied Kendall's tau and Goodman and Kruskal's gamma to contingency table data and constructed confidence intervals to assess the strength of the association of the two syndromes.
- E. Derenzo (CC/Bioethics). The apparent increase in age-related discriminatory language in the popular literature, specifically in skin-care-product advertising, was studied. An LSM quadratic equation model of the age-biased language occurrence rate was used after transforming the data to adjust for the relative increase in advertising pages generally over the years 1969-1988 and stabilizing the observed variances. This model was found to be highly significant in modeling the substantial increase in such language.
- D. Johnson (CC/Bioethics). A lengthy questionnaire survey study was conducted among more than 400 clinic physicians from several of the institutes at NIH to examine the possible differences in the handling of doctor-patient disclosure rates for clinical treatment and progress. Also examined were the possible situational and context-sensitive distinctions in disclosure and the rationales used for the amount of information disclosed, when, and to whom. Attitudinal changes over the last 20 years were studied regarding the amount and timing of physician disclosure to the patient for treatment and prognosis.
- A. Cheever (NIAID/LPD). LSM provided statistical assistance in a study of egg persistence in tissue in schistosome infections after treatment. The purpose was to study the egg production by pairs of worms and their persistence in various body tissue sites in the mouse model (in humans, egg persistence in the liver is associated with serious fibrosis). Statistical methods by LSM included analyses of variance and of covariance. Work on mathematical models for egg accumulation, destruction and evaluation is continuing.
- S. R. Price (NHLBI/LCM). The mitogenic effect of tumor necrosis factor and fetal calf serum on human fibroblasts was studied in six experiments performed at each of several time points in the period of 24 to 48 hours after treatment. Within each experiment, cell growth was compared to untreated cell cultures and mitogenic changes were recorded, but uncontrolled variation across the experiments did not allow pooling of the results. LSM analysis included Fisher's omnibus test for independent tests to generate a summary measure of significance.
- N. E. Rosenthal (NIMH/CPB). In a study of seasonal affective disorder (SAD) in a group of subjects seropositive for human immuno-deficient virus (HIV), the relationship between environmental light entering the eyes and changes in immune function was examined. Eighteen non-SAD and 17 SAD

patients were studied in a two-period cross-over design (without washout phase), with each period consisting of exposure to either low-intensity red light, ambient room light, or high-intensity white (blue spectrum free) light. LSM analysis included maximum likelihood estimation for the effects of changes in immune function, mood scores, and the type and order of light presentation. The EM (Expectation-Maximization) approach to maximum likelihood allowed for a variety of covariance matrix patterns for the measures over time as well as for the effects of covariates.

- J. Grossie (NIAAA). A cluster analysis of evoked potential data was performed to determine whether nerve cells of rats could be categorized according to their response. Six action-potential parameters were used as input to k-means and dendrogram clustering analyses. The results indicated that the 103 cells divided into response curve size. This suggested that there are different types of nerve cells involved.
- E. Goldberg (American University). In a study of how mothers' attitudes toward eating and other factors affect their daughters in terms of eating disorders, LSM provided an analysis based on log-linear models for multidimensional contingency data. Some interesting associations were found.
- D. Rabin (NICHD/DEB). Blood samples were obtained from 11 normal pregnant women at 30-minute intervals over 12-hour periods in order to relate the cyclical variation of specific hormones in their blood. LSM provided statistical advice for analyzing the cyclical variation of human corticotropin releasing hormone (HCRH) and its relationship with the release of adrenocorticotropic hormone (ACTH).
- M. Fulham, A. Brunetti (NINDS/DIR/OD). Positron emission tomography (PET) was employed to measure the uptake of glucose in the brains of 16 patients with Cushing's disease and 16 normal subjects. Abnormal utilization patterns have been observed in several psychiatric disorders. Glucocorticoids affect glucose metabolism and psychiatric symptoms are frequently associated with increased plasma glucocortical levels, such as in Cushing's disease. An LSM analysis compared the two groups of subjects for 20 regions of the brain with a Wilcoxon rank sum analysis adjusted in a stepwise fashion by a Bonferonni correction. Correlation assessed the association of these glucose metabolic rates with clinical and laboratory variables in patients with Cushing's disease. PET scan data were also analyzed for patients with brain tumors. An LSM analysis used a stepwise nonparametric procedure based on Jonckheere's test to investigate which regions of the brain showed a monotone association between metabolic rates and use of steroids (none, less than 1 month, more than 1 month). Of interest has been whether glucose uptake can be predicted by clinical variables, especially severity of Cushing's diseases and psychosis.
- R. Coxey, E. J. Blanchette-Mackie (NIDDK/LCDB). Freeze-fracture cytochemistry and electron microscopy were used to study the accumulation of unesterified cholesterol in the Golgi complex of the fibroblasts from patients with Niemann-Pick Type C (NP-C) disease and those of normal subjects. The hypothesis was that, since the translocation of low density lipoprotein (LDL) unesterified cholesterol from lysosome to plasma membrane is defective in NP-C fibroblasts, such cholesterol is processed in the Golgi compartments differently in the cells of NP-C patients than in those of normal subjects. Cells were either cholesterol-depleted or -repleted. LSM provided a nonparametric analysis based on the Kruskal-Wallis test that compared the densities (in complexes per square micron) in the cis and trans Golgi cisternae and trans Golgi vacuoles.

Research and Development in FY91

Research projects in LSM varied widely; they included statistical methodologies for biomedical applications, scientific computer printing and graphics, research in mathematical and statistical methods, and language processing for medical information systems.

During FY91, staff members of the Statistical Methodology Section were active in mathematical statistical research in a variety of areas with applications to biomedicine. A comprehensive study of Receiver Operating Characteristic (ROC) curves that began in FY88 was continued. Studies of size and shape variables continued to provide methods for studying random proportions or ratios of common occurrence in biomedical data. Studies of patterned covariance matrices continued. Mathematical methods were utilized to investigate problems of statistical inference in the presence of quantum uncertainty.

During FY91, the staff of the Biomathematics and Computer Science Section spent considerable time upgrading the local area network facilities available to LSM and LAS staff. The current configuration includes five 386-based PCs, four 286-based PCs, five Macintosh IIs, and four 3Com3Station microcomputers connected to two LSM/LAS network servers, as well as one Sun 3/50 workstation and one Sun SPARCstation connected to the Advanced Laboratory Workstation local area network. A contract for testing commercial software for mathematics was completed. Biomathematics and Computer Science Section research included development of a methodology for relations in algebraic structures and for new systems of corresponding dual relations. Potential applications to computer methodology and discrete mathematics as applied to biomedical research are under study.

In the Medical Information Science Section, research in medical linguistics continued. Work on the software platform for collaborative lexicographic research was carried forward. Neural networks were investigated for their ability to solve a classical linguistic problem of ambiguity resolution by context. A translator was written for tree-structured NCBI genetic data in ASN.1 format to convert it to Prolog expressions.

Future Plans: Software Development, Contracts, Network Development

LSM's high level of support for MVS-based System 370 statistical software systems will be continued, as will support of DNAdraw, SPU, FPU, and Mathematica. Expanded support of statistical and mathematical software is anticipated on MS-DOS, Macintosh and workstation platforms.

LSM will continue its contribution to software testing and development for two DCRT local area computer networks: a network supporting MS-DOS and Macintosh microcomputers, and the advanced laboratory workstation project of CSL supporting Unix-based workstations. Another server, several Macintosh II microcomputers, and many new or upgraded network peripherals are planned for the former. Support of SAS in a PC/networking environment will also continue. An upgrade of MS-DOS machines to the latest version of the Disk Operating System (DOS) is planned as is the evaluation and implementation of a new network operating system.

Research Projects

Nonparametric Statistics

G. Campbell, Ph.D.
with M.V. Ratnaparkhi, Ph.D. (Wright State University)

A comprehensive study of the theory and application to biomedical research of Receiver Operating Characteristic (ROC) curves has continued. ROC

analysis is used to compare two diagnostic tests via ROC area when the data are ordinal categories rather than continuous variables. Also, the theory of ROC curves has been extended to fuzzy data; instead of knowing that a patient is diseased or nondiseased, it is sometimes useful to view the patient's disease state to be a continuous variable between zero and one, where zero indicates normal and one indicates diseased.

A study of the applicability of Lomax distribution to ROC curves has continued. A paper with Dr. Ratnaparkhi on the maximum likelihood estimation of Lomax models and their iterative fit via computer has been submitted.

An important practical area of ROC analysis recently studied is the incorporation of covariate information. A collaboration joint with Drs. J. Norman, D. Levy and J. Bailey is using age and body mass index (habitus) in a linear regression model to improve fuzzy ROC performance of an ECG test in prediction of left ventricular hypertrophy.

Multivariate Statistical Analysis

J.E. Mosimann, Ph.D.
with G. Campbell, Ph.D. (LSM); M.V. Ratnaparkhi, Ph.D., (Wright State University, Dayton, Ohio); E. Depiereux, Ph.D., (Facultes Notre Dame de la Paix, Namur, Belgium); C. Wiseman (The American University, Washington, D.C.)

The object of this project is the study of multivariate ratios or proportions. Studies of random proportions and ratios, including those that follow a multivariate lognormal distribution as well as those following the distribution of a mixture of Dirichlet distributions, were continued. A paper with Dr. M. V. Ratnaparkhi on normalizing transformations of ratios and proportions using either the beta or unit gamma distributions has shown ranges of values for which proportions may reasonably be transformed to normally distributed random variables so that the wide range of methods based on an assumption of normality may validly be used. Also shown are ranges of values for which it is unreasonable to expect that the usual transformations (logarithmic and Tukey-like) can achieve this. In the past year, a major emphasis has been on comparing the Dirichlet model with small parameters with the analogous lognormal model. Research has continued in the application of ratios in studies of eating disorders such as anorexia nervosa and bulemia. A completed study using size and shape ratios (with C. Wiseman et al) has shown that cultural desirability for thinness in U.S. females has increased in recent years and suggested that this drive for thinness may well be associated with increased prevalence of eating disorders.

Statistical research has begun in the area of statistical forensics in the development of methods for detecting nonrandom patterns in biomedical data. In particular, novel models relying on mixture of Poisson distubutions have been developed and applied. Research has begun on the distribution of nonsignificant numerical digits under various assumptions.

Algebraic Methods in Statistics

J.D. Malley, Ph.D.

Simpler, unified methods for mixed model analysis of variance have been treated. New complete generalizations of the standard tests for statistical independence of sums of squares, given an arbitrary family of allowable, unknown covariance matrices, have been developed. A solution to several forms of mixed model analysis, given missing data, has been obtained using an algebraic solution to the convergence problem for the well-known EM algorithm. This allowed for arbitrary patterns of missing data, as well as for arbitrary forms of the unknown covariance matrix. Using the new methods, convergence is guaranteed for many of

the most common missing-data problems, and applications can be made to longitudinal and growth model problems, as well as to certain genetics studies and time-series problems. The computational power of several well-known statistical methods may be due in part to these new algebraic methods.

A research monograph of these results has been accepted for publication by Springer-Verlag, in the Lecture Notes in Statistics Series.

Quantum Statistical Inference

J.D. Malley, Ph.D.
with J. Hornstein, Ph.D. (Naval Research Laboratory)

This is an interdisciplinary project that investigates statistical inference and estimation for data that obey the rules of quantum mechanics. There are many rapidly developing areas of biotechnology based either on quantum optics, or on technologies that could benefit from a quantum theory treatment. These include: PET scans; real-time, laser-based confocal microscopy of living cells and tissue; bioluminescent molecular tagging; enhanced chemiluminescence; optical communications, fiber-optic sensors, and remote sensing.

The statistical, mathematical and quantum theory machinery needed is intensive, but does permit a transposition of many of the classical statistical methods into these new areas. A manuscript is currently under revision for an article in Statistical Science.

Statistical Inference and Neural Networks

J.D. Malley, Ph.D.

This project has been investigating the possibility of using the methods of neural networks, simulated annealing, and the Boltzmann learning machine for improved computational power for statistical inference and estimation. A recently obtained result proved that the well-known EM algorithm is equivalent to the operation of that form of a neural network known as a Boltzmann learning machine. More generally, the established method of maximum likelihood estimation is known to be realizable as a simulated annealing procedure.

This connection of classical statistical methods and massively parallel computation presages greatly improved estimation with difficult missing value problems. In addition, there exist other, closely-related, recently devised computational methods, such as Gibbs sampling and data augmentation, that are also promising for parallel implementation. These methods share the EM algorithm strategy of enlarging or sampling from an expanded likelihood, with the similar aim of reducing the computational burdens of statistical problems in biomedical research that are otherwise analytically intractable. These problems include repeated measures designs, missing-data problems and constrained covariance estimation.

Computer Graphics and Applications

M.B. Shapiro, M.A.

The main objective of this project is the application of computer graphics and related methods to NIH research problems.

Major updates were made to the DNAdraw program in response to user needs. Version 3.0 allowed for automatic highlighting of aligned sequence data, the user indicating only the number of matches required in a column and the type of highlighting. Version 3.5 provided PostScript output in addition to the normal HP LaserJet output. Also, the manual was completely rewritten and made available as a PostScript file.

In addition, progress was made in creating a special version of DNAdraw for mainframe computers.

Instead of using the menuing features of microcomputers, the user will prepare a template indicating where and what the highlights are.

Discrete Mathematics and Applications

G. Hutchinson, Ph.D.

The project objective is to develop mathematical theory and computational techniques using discrete mathematics (algebra, combinatorics and graph theory), and to apply such methods to appropriate problems of biomedical research and computer science.

Work continued on a category approach toward the definition of algebraic relations. It has been proved that various universal algebraic properties can be treated in a unified way by considering the category regularity conditions for relation and dual relation construction. This included algebraic varieties such as Boolean algebra and modules over a fixed ring. A manuscript "Relation categories and coproduct congruence categories in universal algebra" was completed and submitted for publication. Other research devised formulas corresponding to two-sided ring ideals of rings with prime power characteristic.

Production versions of two LSM-developed public-domain-software products have been distributed. Scientific Printing Utility (SPU, formerly VMAP) provides symbol and font options for scientific text, and File Printing Utility (FPU) offers easy formatting and fonts for general printing on MS-DOS computers. The final research and development of these printing utilities has been completed.

Automated Data Processing of Medical Language

G. Dunham; S. Harper (LSM)
with E. Jaffe (NCI/DCBD/LP), D.E. Henson, M.D. (NCI/DCPC), A. Moshell (NIAMS/EP/SDB)

The major objective of the project is the development of methods for the automatic processing of natural medical language. Research in medical linguistics includes the development of paraphrasing rules for medical noun phrases, a semantic grammar, and a generalized morphological analysis approach suitable for general language and data compression as well as medical language morphology. Development of the Lexicographic Environment Software, a platform for medical lexicography and medical language processing, has continued.

Work has begun on a study of error back-propagation in neural networks to determine what changes would be necessary to allow facile solution of harder problems such as ambiguity of input symbols and the exclusive-or problem.

Collaboration continued on the Clinical Information Utility with the Laboratory of Pathology (NCI/DCBD/LP) and with DCRT/DMB, in using LSM's automatic encoding system to maintain and improve the database of Clinical Center surgical pathology reports. This year, it has been possible to update regularly the encoded database.

Efforts were made with Dr. D. E. Henson to identify linguistic regularities in 296 pathology reports by 8 dermatopathologists engaged in a revision of the terminology for borderline malignant melanoma as part of the preparation for an NIH Consensus Development Conference on the Diagnosis and Treatment of Early Melanoma.

Collaboration has begun with Dr. G. Michaels concerning the rapid prototyping of molecular biology research aids. A translator (in the C language) has been written to convert the *E. coli* component from the NCBI prototypical augmentation of GenBank from ASN.1 format into Prolog. Several databases have been created from this for various collaborators.

Publications

Bachi A, Dunham G, Ginsburg A, Hagstrom R, Joerg D, Kazic T, Matsuda H, Michaels G, Overbeek R, Rudd K, Smith C, Taylor R, Yoshida K, Zawada D. An integrated database to support research on *Escherichia coli*, *Argonne National Laboratory Report* (in press).

Bayne K, Mansfield H, Dexter S, Campbell G, Yamada F, Suomi S. The reduction of abnormal behaviors in individually housed Rhesus monkeys (*Macaca mulatta*) with a foraging/grooming board, 1991;23:23-35.

Campbell G, Levy D, Bailey JJ. Bootstrap comparison of fuzzy ROC curves for ECG-LVH algorithms using data from the Framingham Heart Study. In Bailey JJ, Klingfield P, eds. Computerized Interpretation of the Electrocardiogram XIV, *J Electrocardiol* 1990;23:S132-137.

Cropp CS, Lidereau R, Campbell G, Champene MH, Callahan R. Loss of heterozygosity on chromosome 17 and 18 in breast cancer carcinoma, *Proc Natl Acad Sci* 1990;87:7737-7741.

Del eo JM, Campbell G. The fuzzy receiver operating characteristic function and medical decisions with uncertainty. Proceedings of the First International Symposium on Uncertainty Modeling and Analysis. IEEE Computer Society Press, 1990;694-699.

DeRenzo F, Malley JD. Increasing use of ageist language in skin care product advertising, *J Women and Aging* 1991 (accepted).

Flack M, Oldfield E, Cutler G, Zweig M, Malley J, Chrousos G, Loriaux D, Nieman L. Urine free cortisol versus 17-hydroxysteroid excretion in the high dose dexamethasone suppression test for the differential diagnosis of Cushing's syndrome. *Ann Intern Med*, 1991 (in press).

Malley J. Statistical applications of Jordan algebras. In: "Lecture Notes in Statistics" series, Springer-Verlag, NY, 1992 (accepted).

Rabin DS, Schmidt PJ, Campbell G, Gold PW, Jensvold M, Rubinow DR, Chrousos GP. Hypothalamic-pituitary-adrenal function in patients with the premenstrual syndrome, *J Clin Endocrinol Metab* 1990;71:1158-1162.

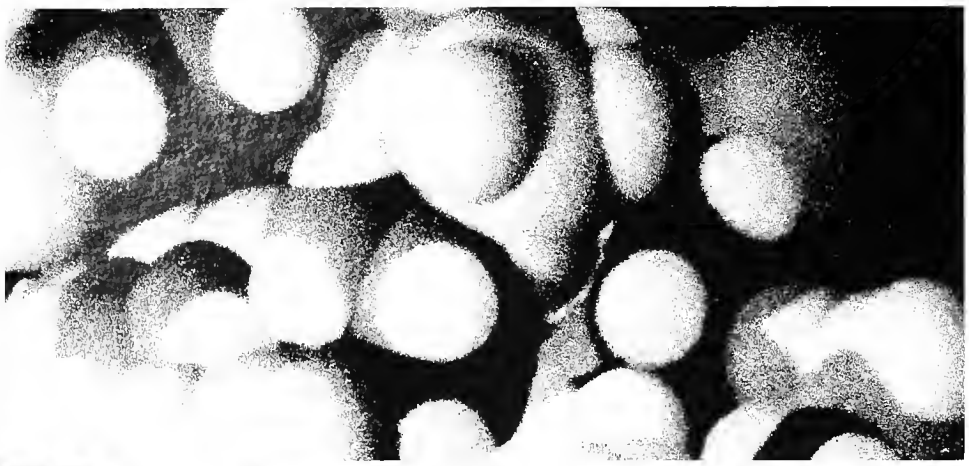
Ratnaparkhi MV, Mosimann JE. On the normality of transformed beta and unit-gamma random variables. *Communications in Statistics Theory and Methods* 1990;19:3833-3854.

Shapiro M. DNAdraw - a program for formatting and drawing DNA sequences, BINARY - *Computing in Microbiology* 1990;20:187-190.

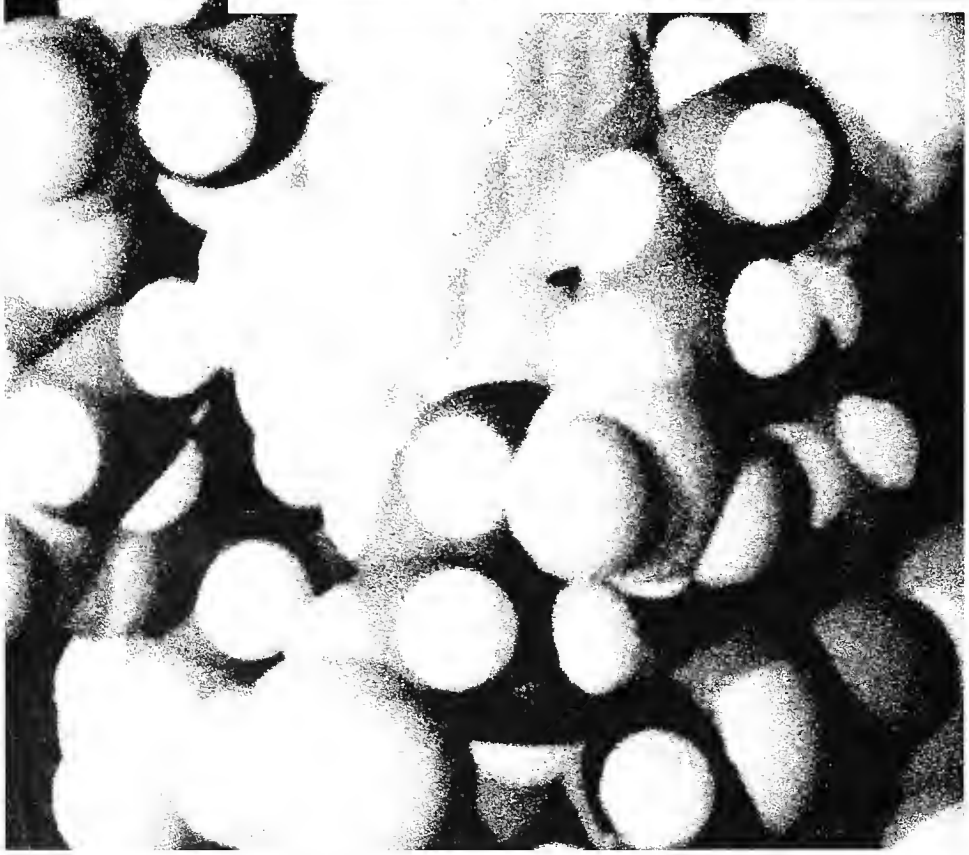
Turkeltaub PC, Campbell G, Mosimann, JE. Comparative safety and efficacy of short ragweed extracts differing in potency and composition in the treatment of Fall hay fever, *Allergy* 1990;45:528-546.

Wiseman C, Gray J, Mosimann J, Ahrens A. Cultural expectations of thinness in women - an update, *Int J Eating Disorders*, 1991 (in press).

Yanovski S, Yanovski J, Malley J, Brown R, Balaban D. Telephone decision making by primary care physicians, *Pediatr*, 1991 (in press).



LAS



Laboratory of Applied Studies

John E. Fletcher, Ph.D., Acting Chief

Executive Summary

In FY91, the Laboratory of Applied Studies' (LAS) efforts were directed toward continued DCRT support for mathematical modeling and data analysis tools such as MLAB and MATLAB. Biomedical signal analysis projects currently use new conditional probabilistic methods in epidemiological studies and also involve heart rate variability, including a new project to study sudden cardiac death by ambulatory ECGs. LAS has a major role in planning and developing the concept of an Imaging Facility into a working prototype. LAS is also helping in the effort to develop a Scientific Computing Resource Center that would provide a single point of contact for intramural scientific computation, especially in the areas of mathematical modeling, data analysis, and biomedical imaging.

The Applied Mathematics Section (AMS) led an effort to continue the MLAB functions to the NIH intramural research program. A bulk purchase contract was negotiated for PC-MLAB. Lead users were identified, trained, and supplied with DCRT-purchased software, including utility programs. The AMS has also developed model-fitting algorithms and stiff differential equations solvers for the MATLAB analysis software. These enhancements provide an alternative system to PC-MLAB that can run on PCs as well as on platforms such as VAX, Convex, and other machines.

The Medical Applications Section (MAS) continued investigations into the prognostic power of the ECG utilizing data from the Framingham Heart Study. A new set of projects involves signal processing methods for nonbiased suppression of low-frequency noise in the ECG, accurate automatic R-wave detection, classification of abnormal beats, coherent signal averaging, and construction of heart rate variability power spectral density. Transfer of signal processing algorithms to the Macintosh system is complete.

The mission of the Laboratory of Applied Studies (LAS) is to provide a collaborative link between DCRT intellectual and physical resources and the NIH intramural research program (IRP) in both basic and applied research. LAS utilizes a multidisciplinary approach to fulfill its mission requiring knowledge and expertise in mathematical modeling, applied mathematics, computer science, and engineering technology. The LAS maintains a balanced approach to research involving both direct collaboration and independent studies of numerical methods and mathematical algorithms applicable to biomedical research to effectively fulfill its mission.

The LAS is divided administratively into two sections: the *Medical Applications Section (MAS)*, staffed by physician-scientists, engineers, and computer systems analysts; and the *Applied Mathematics Section (AMS)*, staffed by mathematicians and computer scientists specializing in biomathematics and bioengineering. LAS staff also function as task leaders and members of DCRT-wide or NIH-wide working groups on IRP projects requiring a multidisciplinary approach.

Intramural Research Support and Collaboration

Fiscal Year 1991, as in previous years, found LAS involved in projects and research support in almost every ICD at NIH as well as in several external collaborations. Specifically, efforts were directed toward continued DCRT support for mathematical modeling and data analysis tools such as MLAB and MATLAB. Biomedical signal analysis projects currently use new conditional probabilistic methods in epidemiological studies and also involve heart rate variability, including a new project to study sudden cardiac events by ambulatory ECGs. Progress was made on several fronts

in the analysis and processing of biomedical images. LAS has a major role in planning and developing the concept of an Imaging Center into a working prototype. LAS is also helping in the effort to develop a Scientific Computing Resource Center that would provide a single point of contact for IRP scientific computation, especially in the areas of mathematical modeling, data analysis, and biomedical imaging.

As in past years, LAS staff were active in various professional societies, meetings, presentations, and other technical events. Staff involvement also included division-wide administrative, planning and managerial functions and representation on various councils dealing with social and equal employment opportunity questions.

Applied Mathematics Section

Fiscal Year 1991 efforts in the AMS were redirected toward software development and implementations of user-friendly data analysis and modeling tools. This redirection was required, in part, because of the release of the DEC-10 from NIH scientific use. The AMS led an effort to restore the MLAB functions to the NIH intramural research program. A bulk purchase contract was negotiated for PC-MLAB. Lead users were identified, trained, and supplied with DCRT-purchased software, including utility programs. The AMS has also developed model-fitting algorithms and stiff differential equations solvers for the MATLAB analysis software. These enhancements provide an alternative system to PC-MLAB that can run on PCs as well as on platforms such as VAX, Convex, or other machines.

The Partial Differential Equations Programs (PDEPGMS) were revised to a standard input format. The memory limitations of 640K imposed by DOS proved restrictive for the PDE Collocation software (PDECOL), and completely inhibitory for the two-dimensional version (PDETWO). The latter was installed and tested on the 3090 mainframe and is available for use. A parallel effort has investigated visualization methods for incorporation with the differential equations software. We are investigating the Quinn-Curtis engineering graphics routines written in FORTRAN as a basis for surface (3-D) visualization. These routines are very fast, but they must be modified and adapted to each particular application. We are currently designing a general-purpose interface for the use of these routines.

The details of the research collaborations with intramural scientists in FY91 are discussed in the individual progress reports.

Medical Applications Section

Medical Applications Section investigations into the prognostic power of the ECG continued with the Framingham Heart Study. A new set of projects involve signal processing methods for nonbiased suppression of low-frequency noise in the ECG, accurate automatic R-wave detection, classification of abnormal beats, coherent signal averaging, and construction of heart rate variability power spectral density. Reactivation of neurosignal analysis activities utilizing evoked potentials is planned. New initiatives for designing quantitative ECG criteria are proposed. Transfer of most signal processing algorithms to the powerful Macintosh system has been completed.

LAS/MAS has continued its strong support of image processing and image analysis through collaboration in Clinical Center and Nuclear Medicine projects. Algorithms for PET and MRI image alignment have been made faster and tests for evaluation of their accuracy have been established. Algorithms for automated PET-to-PET alignment (attenuation scans,

pseudo-attenuation scans, and emission scans) have been written and evaluated.

LAS played a leading role in establishing the functional specifications for an NIH-wide image processing system being developed under a Diagnostic Radiology Research Program contract. LAS is now involved in oversight of the image processing software development by the vendor.

Staff Activities

J. Fletcher (Acting Chief, LAS, and Chief, AMS) is acting as planning coordinator for the DCRT strategic planning activities, both for internal DCRT plans and for NIH-wide planning.

R. Shrager (AMS, LAS) continued to serve as consultant to NIH biochemists in data analysis and experimental design, particularly in areas involving model fitting using MATLAB, MLAB and SVD in multivariate data analyses.

J. Bailey (Chief, MAS) led a group in establishing guidelines, adopted and published by the American Heart Association, for the practice of automated electrocardiography.

Margaret Douglas continues to serve as consultant to NIH intramural scientists in areas of image processing and image software, hardware and systems. She is the LAS representative to the DCRT Scientific Computing Research Center, and is collaborating in the design of a proposed DCRT Center for Image Technology. She chairs the DCRT EEO Advisory Committee and is the DCRT representative to the NIH Advisory Committee for Women's Health Issues.

Future Tools, Trends

In FY92, the conversions from DEC-10 scientific software should be completed, and most IRP investigators will be supported through either PC-MLAB, MATLAB, or an equivalent analysis program.

Future collaborative projects will continue to involve curve-fitting and nonlinear differential equations. More frequently, projects will involve multivariate methods such as singular value decomposition (SVD). The growth of projects involving SVD is due to rapid-scan technology, allowing a large matrix of data to be gathered in a single run. The data analysis technique must therefore be of equivalent capacity and sophistication.

Demands for in-house algorithm and software development should decrease as the quality and quantity of commercial software increases. However, there will continue to be gaps between what is available and the specific needs of IRP projects. Project-oriented software will continue to be an IRP need from DCRT, and this software will be written in very high-level languages like MATLAB.

Other computing platforms, such as the RISC 6000, PS 295, or another scientifically oriented personal workstation, will be considered for use as a laboratory scientific computer. The location and support of this methodology in a Scientific Computing Resource Center will be considered as that facility is equipped and staffed.

A complimentary effort is under way in examining techniques and commercial software for solution visualization. The evaluation of the promising Quinn-Curtis FORTRAN engineering routines is currently in progress. Preliminary results show that this approach is more feasible and can meet the rapid response requirements of online computations. Visualization techniques for solution viewing will be refined and simple menus will be developed for user interaction. PDE solution methods may also be explored on the new DCRT highly parallel computer. New collabora-

tions with IRP members are expected to guide further development and refinement of this software. User training will be offered in the form of seminars and short courses.

A new project will focus upon the prognostic value of heart rate variability (HRV). For that purpose, ambulatory ECG data (AECGs) are being obtained from Harborview Medical Center (University of Washington) and ECG monitoring data from a coronary care unit (Ohio State University).

The AECGs are recorded on special, slow-speed analog media which require automatic correction for wow and flutter. A commercial device that can handle different types of AECG recordings and automatically digitize the data will be loaned to LAS. The data will be transferred to LAS digital media for analysis. ECG monitoring data from the OSU coronary care unit are recorded on analog cassette tapes and will be digitized and analyzed in the LAS facility.

However, the methods for constructing and analyzing HRV spectra are not standardized. LAS will compare a number of methods, including the ways to construct power density spectra of HRV (e.g., FFT, autoregression, polynomial methods), using simulated data of well-defined characteristics. In addition to selecting the best methods for filtering and R-wave detection, LAS will also compare feature extraction methods.

The Framingham study presents an unprecedented opportunity to establish separate ECG criteria for men and women. The echocardiographic, physical and mortality data are supplied by Framingham colleagues. The current hypothesis is that if left ventricular (LV) mass and QRS voltages are adjusted for age and body habitus, the ECG diagnosis of LVH could be significantly improved. Early results support this hypothesis. A more definitive study is being prepared for publication.

Further work is needed to redefine risk factors using a left ventricular hypertrophy (LVH) membership function instead of the dichotomous normal/LVH classification. A method for assigning LVH membership based upon ECG parameters, possibly including ST-T and P-wave parameters as well as adjusted QRS voltage, needs to be developed. The ultimate goal for this work would be to separately determine the association of risk factors to ECG-LVH membership for men and women.

Research Projects

The Solution of Reaction Diffusion Systems in Biology

J.E. Fletcher, Ph.D.
with G. Weiss, Ph.D. (DCRT/PSL); L. Yu, Ph.D. (NIA/LPB); J. Weinstein, M.D., Ph.D. (NCI/DCBD); C. Dong, Ph.D. (NCRR/BEIP)

This project consists of the development of numerical methods and mathematical software for the solution of ordinary and partial differential equations that describe dynamic physiological processes.

In FY91, software to solve coupled systems of two-dimensional reaction-diffusion equations, PDETWO, was installed and tested on the IBM 3090 mainframe. It was also explored using the PS 2/50, and the PS 2/80. All programs were examined with both RM FORTRAN and MICROSOFT FORTRAN. The PDEPGMS software was revised to use the same generic input files for each method, and the methods were adapted for general-purpose applications. The 640KB RAM working space limitations prohibit the solution of large problems with either the PDECOL or PDETWO software. Versions of the PDECOL software were modified for the Convex and are being

applied to the preliminary exploration of dynamic muscle contractile mechanisms.

The manuscript, "A Modeling Analysis of Monoclonal Antibody Percolation Through Tumors: A Binding Site Barrier," coauthored by the principal investigator of this project, was selected for an Award of Merit by the Journal of Nuclear Medicine.

The integration of visual display methods into the solution software was also explored in FY91 (Figure 1, page 80). The purpose of this software is to enable the investigator to view the computed solution of his problem.

Evaluation of the promising Quinn-Curtis FORTRAN subroutines is under way. Preliminary results show that their approach is more feasible and can meet the rapid response requirements of online computations. However, the more esthetic features of graphics such as hidden-line removal, automatic scaling, and origin translation are not included.

Future Tools, Trends

Updated versions of commercial PC-FORTRAN compilers promise to avoid the 640-KKB limitations of DOS through the use of DOS extenders. At least one (RM FORTRAN) is on order for testing. PDE solution methods may also be explored on the new DCRT highly parallel computer. Visualization techniques for solution display will be refined and simple menus will be developed for user interaction. New collaborations with NIH scientists are expected to guide further development and refinement of this software. User training will be offered in the form of seminars and short courses.

Cellular Kinetics Models of the Human Immune System (An Investigation of HIV-like Infections in a Model Immune System and its Response to Opportunistic Pathogens)

J.E. Fletcher, Ph.D.
with J.J. Bailey, M.D., R.I. Shrager, M.S. (DCRT/LAS); H.C. Lane, M.D. (NIAID/DCD); M. Basler, Ph.D. (PRI/FCRC)

LAS is exploring mathematical models of the human immunoregulatory network and the kinetics of its many complex interacting components (i.e., precursors, CD4+ and CD8+ T-cells, B-cells, T-cytotoxic, NK and LAK monocytes, interleukins, and interferons) by means of a system of nonlinearly coupled ordinary differential equations. An appropriately constructed and validated network model should suggest experiments that improve the understanding of how the immune system might be manipulated to increase its effectiveness in preventing or neutralizing pathogenic infections.

In FY91, models of human immunodeficiency virus (HIV) interaction with human CD4+ T-cells *in vitro* were studied to determine their stability characteristics. The experiments being modeled represent healthy and HIV-infected CD4+ T-cells in a culture designed to keep the cells growing. The complex issues in modeling even this simple system are several: direct cell-to-cell transmission of HIV vs. transmission via the supernatant culture media, delay of infectivity, delay of viral synthesis, nutrient limitation, and the effects of a antigenic stimulation and/or repeated stimulations.

Individual CD4+ cell culture experiments were designed and conducted with the cooperation of NIAID investigators and their support contractors at the Frederick Cancer Research Center. Experimental results designed to determine *in vitro* CD4+ and CD8+ culture parameters have been inconsistent.

Figure 1. Perspective views of computed surfaces.

Determinations of life-cycle rates for healthy and infected cultures and viral life-cycle times were not consistent with published literature values. The sources of these differences are now being sought. A manuscript detailing an analytical model with parametric exploration of its characteristics has been developed. The project is in contact with other NIH investigators with regard to experimental data.

Future Tools, Trends

In FY92, our focus will be on refining our parameter values from additional experimental sources, expanding the scope of our primitive model to more than HIV interactions with CD4+ cells. Interaction with other HIV investigators at NIH will continue. Also, we plan to bring in outside investigators to guide and critique our model-building process. The ultimate goal is to progress to experiments that concern antiviral drug treatments, vaccine design and effectiveness, and such areas where modeling may contribute.

Mathematical Modeling Systems

J.E. Fletcher, Ph.D.
with R.I. Shrager, M.S., (DCRT/LAS); G. Campbell, Ph.D., M.W. Hodges, M.S., (DCRT/LSM)

The objective of this project is to acquire, make available, and support mathematical modeling and data analysis software systems that are accessible to investigators from many disciplines. The principals involved develop, test and implement such software on DCRT-supported computing platforms. They develop supplementary software and utilities to optimize the use and efficiency of such software systems. In addition, consultation and training are made available both through formal DCRT-sponsored courses and through individual consultations.

FY91 Progress

The phase-out of the DEC-10 System, which was finalized on February 28, 1991, also resulted in the phase-out of all mainframe versions of the MLAB mathematical modeling system. In anticipation of this phase-out, LAS and LSM personnel led a coordinated effort to save all important MLAB data and work files from the DEC-10 System. The adoption of the Civilized Software, Inc. (CSI) product, PC-MLAB, was recommended as an alternative system. The Advisory Committee on Computer Usage (ACCU) endorsed this recommendation, and a plan for the acquisition, distribution and retraining in the use of PC-MLAB was developed. A group of lead users was identified in each ICD. These investigators were provided DCRT-purchased copies of the PC-MLAB software and low-cost training in its use. The lead investigator in this effort also developed utility software for PC-MLAB that optimizes its available memory under DOS by automatically reconfiguring the PC for MLAB before each use and restoring the previous configuration after the MLAB session. LAS has equipped three PC-workstations with PC-MLAB and LSM has provided a fourth. One workstation in each lab is made available for demonstration, consultation, and for IRP use. LAS and LSM personnel are providing assistance on the conversion of DEC-10 files to PC-MLAB format. They also provide advice and assistance on the development of utility routines and DO files for routine procedures. Manuals, self-teaching guides, and other written guidelines have been produced to facilitate adoption of the new MLAB system by all levels of users.

Future Tools, Trends

LAS and LSM personnel will continue to provide assistance in the purchase, installation and use of PC-MLAB. Some investigations are under way that explore AIX versions of PC-MLAB and consideration will be given to extensions to other computing platforms. These services and support can be provided in the Scientific Computing Resource Center when it obtains appropriate computing platforms. Collaborations and consultations will be provided to the IRP and the ICDs as necessary.

Mathematical and Computational Methods for Solving Nonlinear Equations

R.I. Shrager, M.S.
with J. Fletcher, Ph.D., J.J. Bailey, M.D. (DCRT/LAS); P. McPhie, Ph.D. (NIDDK/LBM); H.C. Lane (NIAID/PIR); R. Berger, Ph.D., R. Hendler, Ph.D. (NHLBI/LC); R. Winslow, M.D., K. Vandegriff, Ph.D., V. McDonald, M.D. (LAIR Blood Res.); A. Alayash, Ph.D. (CBER/NIH); C. Bonaventura, Ph D. (Duke U. Marine Biomed. Ctr.); Z. Dancshazy, Ph.D. (Hungarian Acad. Sci.); P. Berk, M.D. (Mount Sinai Med. Ctr., New York); M. Basler, Ph.D. (PRI/FCRC)

The purpose of this project is to provide NIH investigators with mathematical tools for insight, analysis, and solution of complex equations that arise in the modeling of biological systems. To facilitate these efforts, LAS develops mathematical methods that are accessible to investigators from many disciplines. Software packages that result from these developments are made available to the research community as general research tools. Advice on the use of certain commercial mathematical software packages is also offered.

FY91 Progress Part I: External Collaboration Projects

• Conformational Changes in Hemoglobin (Hb) Binding (with R. Winslow, M.D. and K. Vandegriff, Ph.D., Letterman Army Institute of Research). Hb-oxygen binding is being studied using singular value decomposition (SVD) and the most precise optical spectra that currently available equipment can provide in an effort to detect conformational changes. At least one spectral change has been detected that does not correspond to simple Hb-oxygen saturation (presented at the annual Biophysical Society meeting in February 1991).

• Thermal Unfolding of Swine Pepsinogen (with P. McPhie, Ph.D., NIDDK). A paper on this topic—using SVD and least squares analysis—is in preparation, which concludes that there are at least two stages of denaturation which seem to occur independently, in parallel.

• Oxidation of Heme Groups (with V. McDonald, M.D., LAIR, A. Alayash, Ph.D., CBER, and C. Bonaventura, Ph.D., Duke University). SVD and least squares are being used to explore the possibility that ferrous forms exist in Hb, and that Hb can act as its own reducing agent.

• Relaxation Kinetics of Bacteriorhodopsin (bR) (with R. Hendler, Ph.D. and R. Lozier, Ph.D., NHLBI, and Z. Dancshazy, Ph.D., Hung. Acad. Sci.). The kinetics of bR after-laser flash seems to depend on the intensity of the flash. Explanations for this dependence are being sought using SVD and least squares.

• Microcalorimeter Model (with R. Berger, Ph.D., NHLBI). Differential equations and least squares are being used to improve resolution of heat capacitance and resistance in various solutions.

• ATP-ADP Equilibrium (with R. Berger, Ph.D., NHLBI). ATP, ADP and PO_4 concentrations are

being resolved from near-infrared spectra using partial least squares (PLS). There are problems with noise in the data and with the reliability of PLS itself, yet the early results are promising.

• Hepatic Heme Turnover (with P. Berk, M.D., Mt. Sinai Med. Ctr., New York). Numerical deconvolution and least squares are being used to resolve turnover rates in man and rat to study the effects of phenobarbital on the liver.

Part II: Methodology and Software Development

• Partial Least Squares (PLS). M-files in MATLAB have been written to allow investigators to explore the results of PLS under various assumptions about the rank of the calibration set.

• Stiff Ordinary Differential Equations (SODEs). A new solver has been developed by AMS staff which is especially useful in perturbation studies, where one wants the stability of the solution to match the stability of the system being solved. The new solver has been made available in both MLAB and MATLAB. Projects using SODE models include microcalorimeter simulation and HIV *in vitro* modeling.

• MATLAB. This product has been purchased for the Convex, while MLAB has not, and ML/e is no longer being supported. It appears that MATLAB will become the short-term *de facto* replacement for MLAB on a scientific mainframe, except for some gaps. Some of the gaps have been filled by writing a nonlinear least squares curve-fitter, a stiff differential equation solver, and a multicurve plot routine in the MATLAB language. All are available for general use. A course is being developed, and will be offered when the MATLAB software is available on the Convex.

Future Trends

In FY92, there will be more complete data and models concerning the conformations of hemoglobin and the oxidation states of the heme ion. A new high-speed spectrophotometer will be producing data about the kinetics of the cytochrome chain and bacteriorhodopsin. The SODE model of the microcalorimeter will become a production tool for routine use with the instrument. Evoked potential data from patients in various stages of anesthesia will be available in mid-1992 for analysis by signal processing techniques to see if reliable parameters can be extracted for monitoring the depth of anesthesia.

Computer Systems and Applications for Nuclear Medicine

M. Douglas
with J.J. Bailey, M.D., P. Kalkowski (DCRT/LAS); S.L. Bacharach, Ph.D., M.V. Green (CC); R.O. Bonow, M.D. (NHLBI)

LAS develops systems for computer-based mathematical analysis, pattern recognition and image processing in support of diagnostic activities in the Nuclear Medicine Department of the Clinical Center and collaborating institutes. Many applications are directed toward the correlation of function with structure, such as estimation of ventricular function from radionuclide ventriculography or PET scan (functional data) compared to MRI or CT scans (anatomical data). Other applications are directed to special techniques such as the detection and imaging of tumors using radio-labeled monoclonal antibodies. LAS, in collaboration with the Nuclear Medicine Department of the Clinical Center, has over the past 5 years designed and specified a general-purpose image processing system, MIRAGE. Programming was performed by contractors supervised by LAS and the Nuclear Medicine

Department. The completed basic system has been ported to several other NIH computer systems, including VAX workstations and Macintosh systems.

FY91 Progress

MIRAGE functionality has been included in the design of the new Multimodality Research Image Processing System currently being developed for use across NIH. LAS participates in investigation of the use of neural networks in the problem of image segmentation and classification, campus-wide consultation and collaboration on image processing, especially in the area of PC-based image processing; review of the current state of PC image processing software; and investigation of PC Computer Aided Design (CAD). LAS personnel serve as image processing representatives to the NIH campus. LAS has sponsored seminars through the NIH-wide Image Processing Group, collaborates in publishing the Directory of NIH Image Processing Facilities, and has organized and chaired a 3-D Medical Image Processing Conference. LAS has also assisted in the NCI/NCRR image processing lecture series.

Alignment of MRI and PET images has continued. Previous procedures to align PET and MRI were improved by new subroutines which reduced the time required for alignment and matching.

Studies of NM/PET alignment lead investigation of PET/PET alignments (attenuation scans, pseudo attenuation scans, and emission scans). Even though contaminated with radiation from residual isotope, serial "pseudoattenuation" data can be aligned to the original attenuation data and allow the use of the original attenuation data for correcting subsequent PET emission data. Alignment of PET cardiac image data is a necessary prerequisite to quantitative analysis of serial data acquired on the same patient, or to patient data compared with atlas data. Alignment techniques can correct for patient motion between scans, allowing costly, time-consuming studies to be salvaged and reducing the time a patient must spend immobile during a series of scans. LAS compared the correlation coefficient technique with a sign change criterion to determine translational alignment of PET attenuation images along the x, y, and z axes. The correlation coefficient technique appears to be slower but more accurate when applied to 126*126 (3.5 mm xy resolution) test cases. A method for automated x,y, z axis registration of PET attenuation data has been developed and tested. Progress on the development of registration algorithms was accelerated when LAS acquired a VAX3100 workstation and installed MIRAGE on the system.

Future Tools, Trends

The Nuclear Medicine Department has several systems that collect three-dimensional data. LAS will continue to develop three-dimensional visualization methods, multimodality registration, and fast interactive algorithms for analysis of large volumes of data including rotational alignment, translational alignment, and oblique alignments. The generation of the regions for alignment testing will be automated.

Clinically useful techniques for visualization of volumetric data will be developed, e.g. the simultaneous display of multiple volumes for the visual verification of volumetric alignment. LAS will collaborate in the design of an image transmission network; design of central, short-term image storage and management; and development of a consensus approach to medical image archiving issues.

Multimodality Research Image Processing System

M. *Douglas*
with R. Levin, Ph.D. (OD/DRRP), R. Carson, Ph.D., J. Frank, M.D. (CC), T. Zeffiro, M.D. (NINDS/MNB), G. Sobering, Ph.D. (NCRR/BEIP)

The purpose of this project is to develop an image processing system for the study of two-, three- and four-dimensional data from multiple imaging modalities. The system is to be based on a common hardware and software environment across NIH and is to be the standard system for macroscopic image processing (Positron Emission Tomography, Magnetic Resonance Imaging, Computed Tomography, Single Photon Emission Computed Tomography, echo, etc.) at NIH. The development of a new system that could incorporate all of the functionality of the many old NIH systems plus new 3-D functionality was made possible through funding for the new Diagnostic Radiology Research Program (DRRP).

FY91 Progress

LAS staff, because of their experience with the design of MIRAGE, designed the functional specifications for the image processing software and collaborated in the system design. LAS interviewed key personnel involved with the major image processing systems on campus and investigated several available academic and commercial packages. LAS led the Technical Committee in developing the functional specifications for image processing software which went into the final RFP. LAS staff wrote the image processing portion of the RFP, served as an evaluator of the contract proposals, wrote clarification questions to the vendors, assisted in the business negotiations and designed payment plans, debriefed nonwinning vendors and currently serves as the alternate Project Officer.

Future Tools, Trends

LAS will provide one of the two primary overseers of the contractors in their development of the image processing software. Under the contract, DCRT is purchasing one of the workstations, which will allow LAS to supervise vendor software development and will give LAS a new platform for developing applications for their ongoing collaborative research in nuclear medicine. This new platform will be the first major common hardware and software environment for image processing for use across NIH. Researchers across NIH have already requested over two dozen copies of the software. More than 50 users are expected by the end of FY92.

Although a limited, fixed-price contract for the system was awarded, most research applications will require additional modules. Hence, the system design and specifications emphasize extensibility. There will be ample opportunity for LAS to collaborate in developing additional modules to extend system capabilities.

Computer-Based Analysis and Image Processing in Electron/Light Microscopy and X-Ray and Electron Energy Spectroscopy

M. *Douglas*

The objective of this project is to evaluate computer systems and develop algorithms for mathematical and statistical analysis of image data, pattern recognition, and image processing, principally from two-dimensional micrographic sources such as x-ray micrography and electron energy loss spectra, and from the electron/light microscopy images of biological specimens.

FY91 Progress

During 1991, a Macintosh-based system was established for evaluation of image processing software

applications. A PS/2 system is also used to evaluate application image and graphics software. This year these systems have been used to test such packages as Aries, MedVision, Image, ImagePro Plus, Micro Science Count and Measure, CADD 5.0, Claris CAD, and Transform. These systems were also used as test beds for NIH researchers wishing to test the feasibility of a PC-based image analysis system. LAS staff have developed special-purpose image processing software and procedures which have enabled researchers to use these systems for their investigation of micrographic images, including evaluation of micrographs of intermembrane particles and measurement of compartments of the golgi membrane for analysis of cholesterol transport.

Future Tools, Trends

Future efforts include increased use of the Macintosh platform for image processing, and integration of the Multimodality Research Image Processing System software for microscopy imaging. Both the AT (286-based) and Macintosh systems will be upgraded. LAS expert staff have made these systems available to NIH researchers. This availability will continue through LAS collaboration with the proposed DCRT Scientific Computing Research Center. LAS personnel are working closely with CSL to create a DCRT Center for Computer Image Technology. When such a center is formed, the 286-based PC and Macintosh systems will be included in it.

Computer-Aided Analysis of Electrocardiography

J.J. Bailey, M.D.
with E.W. Pottala, Ph.D. (DCRT/LAS); G. Campbell, Ph.D. (DCRT/LSM); Framingham Heart Study; Field Studies and Biometry Branch, NHLBI; Association for Advancement in Medical Instrumentation

These studies are directed toward evaluation of the prognostic power of the electrocardiogram when analyzed by advanced computer methodology, and the predictive accuracy of diagnostic criteria when implemented in ECG computer programs. Digital signal processing of the electrocardiogram is a problem area requiring considerable engineering and computer science expertise as well as knowledge about its clinical revelance. The use of well-documented populations and multivariate statistical techniques in designing new criteria are also subjects under investigation. Studies have been pursued in collaboration with NHLBI, the Framingham Heart Study, and the European Project for Common Standards in Quantitative Electrocardiology (CSE).

LAS and LSM have developed a number of statistical methods to compare diagnostic criteria and the relative power of ECG diagnostic algorithms. Some of these methods require the use of fuzzy set theory and fuzzy logic, a powerful mathematical technique which has been in use for 15 years.

Left ventricular masses (LVMs) in the Framingham population are distributed from normal to severely abnormal cases. We have defined a "LVH membership" function based upon echo-determined mass. Cases with LVMs not exceeding 1 std above the normal population mean were assigned "zero" membership and cases with LVMs at least 4 stds above were assigned membership "one." For the intermediate cases (from 1 to 4 stds above normal mean) the LVH membership function was assigned in a linear fashion (from 0 to 1 as LVM increased). In this way, a quasi-continuous LVH membership replaced the former discrete categorization (viz. normal or LVH). This allowed construction of "fuzzy ROC" curves for comparing ECG diagnostic algorithms.

FY91 Progress

The method for analyzing the statistics of fuzzy ROCs was used to show that adjusting QRS voltage for age and body habitus could significantly improve ECG diagnosis of LVH (presented at the 16th annual conference of the International Society for Computerized Electrocardiology). Adjustment formulae were derived by regression analysis on the Framingham population, producing different coefficients for males and females.

Future Tools, Trends

Further work is needed to redefine risk factors using an LVH membership function instead of the dichotomous normal/LVH classification. A method for assigning LVH membership based upon ECG parameters, including ST-T and P-wave parameters as well as adjusted QRS voltage needs to be developed. The ultimate goal for this work would be to determine the association of risk factors to ECG-LVH membership for men and women separately.

LAS plans to investigate a method for quantitative ECG criteria using fuzzy set theory and fuzzy logic. Such quantitative criteria would be more stable and less complex than currently used criteria. The fuzzy logic method differs also from the usual probabilistic approach in that major diagnoses are not treated as if they are mutually exclusive. Separate ECG criteria derived from the age and body habitus adjustment formulae for women is an issue which has not been previously addressed.

Analysis of Physiological Signals

E.W. Pottala, Ph.D.
with J.J. Bailey, M.D. (DCRT/LAS); J.A. Dvorak, Ph.D. (NIAID/LPD); K. Rasmussen, Ph.D. (NICHD/LCE); W.C. van Arsdel, Ph.D. (FDA); M. Cowan, Ph.D. (University of Washington); C.F. Baker, Ph.D. (Ohio State University); R.A. Franker, B.Sc. (Medical College of Ohio); R.W. Bowser, B.Sc. (Creighton University); E.C. Phoebus, Ph.D. (University of Puerto Rico)

This project involves the development and application of microcomputer-based signal processing techniques for analysis of physiological signals, e.g., electrocardiogram, electromyogram, and electroencephalogram. The LAS microcomputer-based systems provide a general-purpose analog-to-digital conversion facility and an ability to filter the signals with a variety of analog and digital techniques.

The LAS, in collaboration with NIAID, developed a reliable mouse model for Chagas' myocarditis. Analysis of ECGs allowed each animal to be compared with itself and also indicated which animals would show significant pathology. LAS has completed the processing of the mouse ECGs for the NIAID myocarditis model. A definitive report on the correlation of ECG changes with histopathological lesions of the myocardium is forthcoming pending a review of the tissue sections by the collaborating pathologist.

The collaboration of LAS and FDA provides an opportunity to develop more sensitive but reliable ECG indices of cardiotoxicity by automated methods and thereby set a new standard for therapeutic indices.

Investigators at the Laboratory of Comparative Ethology (LCE, NICHD) and the Psychology Department of the University of Puerto Rico are using the ambulatory ECG to study physiological correlates of behavior in a simian colony. Studies of behavioral physiology in simians may reveal similar traits in humans and indicate ways in which patients can lower their cardiac risks by changing their lifestyle. This work is closely related to a new LAS project on human ambulatory ECGs, and we are designing a PC-based method for analog to digital conversion, filtering, R-wave detection, and construction of heart rate variability power spectrum.

FY91 Progress

LAS has developed a single-pass filter to suppress baseline wander, which will not seriously alter the ST segment and which will operate in near real-time. This will be important for ECG monitors in intensive care units where filters, which suppress baseline wander, seriously distort the ST segment, resulting in false diagnoses of myocardial ischemia or injury.

In FY91, most signal processing activities were transferred to the Macintosh II system. This system can not only perform high-speed analog-to-digital conversion, but also perform any additional signal processing that is necessary. Special Macintosh applications programs have been developed in the C language to generalize and simplify the use of the array processing board.

Our work with the rodent ECGs provided an opportunity to develop an index of QRS morphology based upon Tchebychev polynomials which was computationally simple and fast. This index was useful for beat matching and signal averaging of QRS patterns; it was also useful for excluding abnormal QRS waves from the averaging process.

Future Tools, Trends

LAS has embarked upon a project involving the prognostic value of heart rate variability (HRV). Data for this study in the form of ambulatory ECGs (AECGs) are being obtained from Harborview Medical Center (University of Washington); and ECG monitoring data are being obtained from a coronary care unit at Ohio State University.

The AECGs are recorded on special, slow-speed analog media that require automatic correction for "wow" and "flutter." A commercial device that can handle different types of AECG recordings and automatically digitize the data is being loaned to LAS. The raw digital data will be transferred to LAS digital media for analysis. ECG monitoring data from the coronary care unit (Ohio State University) are recorded on analog cassette tape and will be digitized in the LAS facility and again analyzed by LAS algorithms.

The methods for constructing and analyzing HRV spectra are not standardized. LAS will compare a number of methods (e.g., FFT, autoregression, polynomial methods) using simulated data of well-defined characteristics. In addition to selecting the best method for filtering and R-wave detection, LAS will compare feature extraction methods. We plan to use our Tchebychev-based QRS index to remove abnormal beats from the HRV spectra. Data adjustment for subtracted beats will require further study.

To validate this methodology, we are constructing an ECG simulation that employs an integral frequency modulation model (IFMM) as the underlying generator for RR intervals. The model will build in HRV modulating frequencies and random noise. The output of the IFMM gives timing for a stream of triangular QRS-like pulses. This data stream simulates ambulatory or intensive care monitoring ECGs and can be used to test the filtering and R-wave detection methods as well as the procedure for constructing the HRV spectra.

Seminars and Conferences

LAS supported the following seminars and conferences:

- *Mapping the Human Brain: New Images, New Directions*, Este Armstrong, Ph.D., Nov. 16, 1990
- *Optical (Infrared) Functional Neuro-Imaging in Real Time*, Alexandr M. Gorbach, Ph.D., Eugeni N. Tsicalov, Ph.D., Dec. 13, 1990
- *Image-Processing Applications of Associative Neural Networks*, Joseph Schmitt, Ph.D., Feb. 12, 1991

• *Indexing Electronic Image Databases Using a Geometric Reasoner Hemant*, D. Tagare, Ph.D., May 9, 1991

• *Visualization in Historical Context - Computer Images, Visual Thinkers, and Mental Models*, Thomas G. West, May 22, 1991

• *Imaging Functional Anatomy*, Karl Friston, M.A., MRCPsych., May 30, 1991

• *Chaos and the Electrocardiogram: Fact and Fiction*, David E. Albert, M.D., June 19, 1991

• *Automated Processing of Brain Images Obtained with Positron Tomography*, Mark A. Mintun, M.D., Aug. 28, 1991

• *A Lie Group Theoretic Approach to the Invariance Problem in Feature Extraction and Object Recognition*, James B. Cole, Ph.D., Sept. 17, 1991

• *3D Medical Imaging - A Discussion of the Current State at NIH and Future Directions*, Stephen Bacharach, Ph.D., Richard Coppola, Sc.D., Joseph Frank, M.D., Michael Green, Barry Horwitz, Ph.D., Daniel Rio, Ph.D, Geoff Sobering, Ph.D., Thomas Zeffiro, M.D., Sandra Zink, Ph.D., May 17, 1991

Publications and Presentations

Barest GD, Bacharach SL, Douglas MA, Nhsovec A. Myocardial wall thickening from gated NMR images: rethinking the meaning of thickness, *Proceedings of Computers in Cardiology* 1991;249-252, Venice, Italy.

Campbell G, Norman JE, Levy D, Bailey JJ. Age and habitus adjustment of ECG criteria improve detection of LVH as shown by fuzzy ROC curves, *J Electrocardiol* 1991;23, (in press).

Douglas MA, Bacharach SL, Kalkowski PJ. Alignment of PET attenuation images, *Proceedings of Computers in Cardiology* 1991;629-632, Venice, Italy.

Fletcher JE. *An introduction to PC-MLAB* (a tutorial manual), June 1991, AMS/LAS/DCRT, 50 pp.

Frankel R, Pottala EW, Bowser RW, Bailey JJ. A filter to suppress ECG baseline wander and preserve ST segment accuracy in a real time environment, *J Electrocardiol* 1991;24:315-23.

Hakim Din M, Fletcher JE. Software modification of the Quinn-Curtis graphics routines for PDEPGMS applications, 1991, LAS/DCRT/NIH.

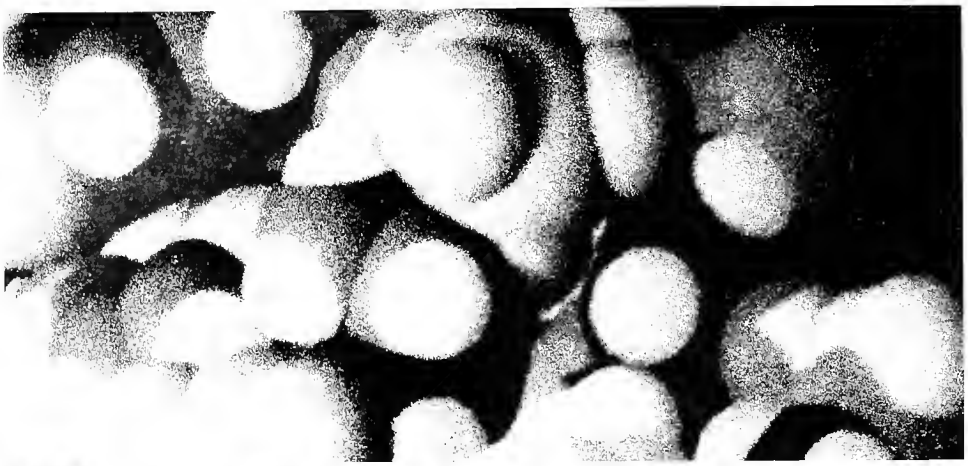
Okuda H, Tavolomi N, Blaschke TF, Kiang C, Jones MJT, Waggoner JG, Sardina MK, Sassa S, Shrager RI, Berk PD. Failure of phenobarbital to increase early labeled bilirubin formation from 4-[14C]-alpha-aminolevulinic acid in man and rat, *Hepatology* 1991 (in press).

Pottala EW, Bailey J. Digital filters to suppress ECG baseline wander, Association for the Advancement of Medical Instrumentation 26th Annual Meeting, May 1991

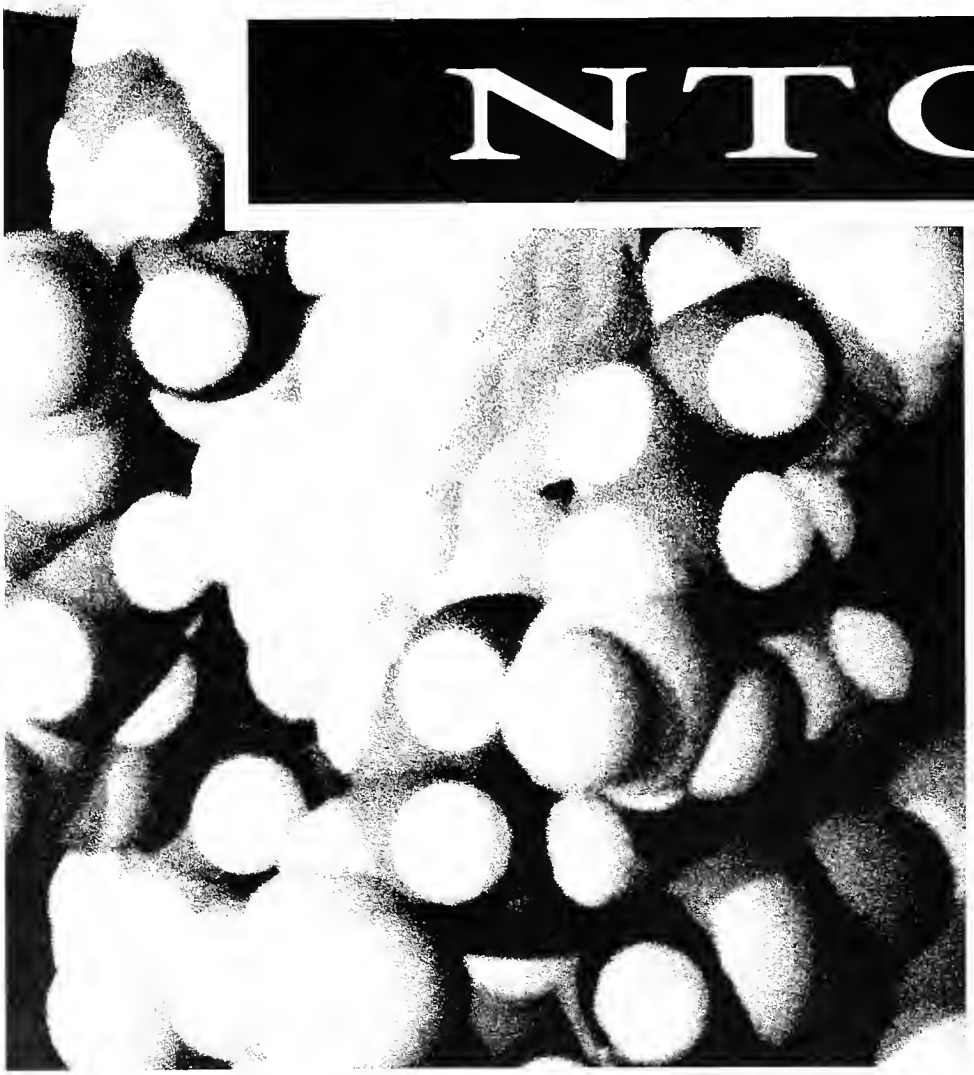
Pottala EW, Bailey JJ, Fletcher JE. A simulation of T-lymphocyte interactions with Human Immunodeficiency Virus. In: Anderson, JG, ed. Proceedings 1992 *Simulation in Health Care and Social Services Conference* 1992;73-77, Newport Beach, California.

Tyler EC, Horton MR, Krause PR. A review for algorithms of molecular sequence comparison. *Comput Biomed Res* 1991;24(4):72-96.

Vandegriff KD, LeTellier YC, Shrager RI, Winslow RM. Analysis by singular value decomposition of the visible spectral transition of oxygen binding to hemoglobin, Biophysical Society 35th Annual Meeting, San Francisco, February 24-28 1991.



NTG



Network Task Group

David C. Songco, Chief

The Network Task Group (NTG) develops and applies state-of-the-art computer communication technologies to support and enhance NIH's biomedical and clinical research programs. The NTG was established in May 1989 to emphasize DCRT's networking activities in three areas:

- design and implementation of a backbone network to link local area networks (LANs) on campus
- guidance and support for locally managed networks
- design and support for DCRT networks.

NTG consists of a group of six engineers who track technology and market trends, perform equipment studies, and design evaluations to provide the most reliable data transport system to the NIH scientific community.

Background

Networking is critical to DCRT's role in support of computing at NIH. Two complementary network backbone development efforts exist within DCRT: NUnet and RESnet. NUnet provides NIHwide connectivity with T1 technology and operates at 1.5 million bits per second. RESnet operates at 10 to 100 million bits per second for more demanding scientific applications, and is implemented by NTG.

RESnet

Currently, RESnet operates using two proprietary 10 million bits per second Ethernet over-fiber-technologies. It spans nine buildings on the NIH campus: 10, 12, 12A, 13, 29, 29A, 30, 36 and 37.

The goal of RESnet has always been to provide the NIH scientific community with a backbone to support applications requiring high-speed communication. From the outset the design goal was to use standardized technology so as not to get locked into any single-vendor solutions. For these two reasons, NTG began evaluating the 100 million bits per second FDDI in September 1990. The evaluation consisted of testing the equipment's compliance to the FDDI standard, the capability of routing the DCRT-supported protocols over FDDI, and the capability of using both multimode and singlemode fiber optic cable as the transmission medium.

NTG has completed the first phase of its evaluation of the FDDI implementation and has found it to be reliable enough to begin the migration process of RESnet to FDDI. The migration process is currently under way, with four of the RESnet routers implementing FDDI. By September 1991, all 10 of the RESnet routers will be running FDDI, and NTG will be looking to expand the FDDI technology throughout the NIH campus.

As new higher speed standards evolve, their progression will be continuously tracked to keep up with changing technology so that the communication speed and service required by members of the NIH community can be provided.

Closet Space

One of the more frustrating and difficult tasks has been locating and gaining possession of physical closet space. The importance of having secure network closets was recognized at the onset of the project. Physical security of the equipment is required for reliable operation of a network. Therefore, an effort to establish network closets in each building was initiated. Over the course of several years, space was finally allocated in every building that the network will service.

We presently have 10 interbuilding closets completed and in use. An additional 28 are in various stages of the procurement/construction process. In addition, we have a total of 38 riser closets in highrise buildings (located in Buildings 31, 36 and 37) which have been constructed to date. The riser closets for

Building 10 (Clinical Center) are in the design phase. A total of 39 riser closets are planned for the Clinical Center.

Management

NTG was chartered to evaluate and implement the best mix of networking technology in support of scientific computing at NIH. A major part of this task is the development of management technology and processes to provide the NIH community with uninterrupted network service. To become more responsive to the user community, we feel it is necessary to automate many of the management tasks that are required by RESnet. This frees the members of the group for manpower-intensive tasks such as

- problem identification and resolution
- improving and upgrading the existing network
- periodic maintenance of the network
- planning for the expansion of the network
- monitoring current trends in technology and discovering how those trends might be applied to the current or to future networks on the NIH campus
- automating more of the network management functions.

Currently, NTG has implemented the Cisco NetCentral management package, and plans to evaluate other network management packages are being made. Our requirements of these management packages are not trivial, and the evaluation may become a lengthy process. NTG plans to evaluate the commercially available packages which seem appealing from both a features and a cost point of view.

Documentation

Documentation of RESnet is an important task. We are currently moving to a single, standard set of documents that will be maintained centrally. This standardized set of documents will begin with the notes that each staff member has kept. The documents will also draw on information that has been provided by subcontractors and by the NIH Division of Engineering Services.

To generate the needed documents, NTG chose Autocad, a commercially available software package, because it could be run on our existing Unix workstations and is a *de facto* standard for architectural and engineering drawings. Thus, the staff could easily share drawings with the Division of Engineering Services and subcontractors. In addition, Autocad is networkable, allowing members of the NTG access to drawings via the network from their individual workstations.

Networking the Clinical Center

The NTG has been working closely with the Clinical Center staff since November 1989 to design and install a backbone network within Building 10. Since it is so large, the Clinical Center network (CCnet) can be considered as a separate campus backbone. Eventually, CCnet will be a comprehensive cabling infrastructure servicing the entire Clinical Center with a mixture of fiber-optic and unshielded twisted-pair cable. An interim infrastructure currently exists in Building 10 and has been in use since February 1990. Over 30 Ethernets are connected via CCnet and run over a variety of media.

Three LAN closets will be built on each of floors 1-13 for a total of 39 closets. Space for these closets has already been allocated. NTG has contracted with an architect and an engineering design firm through the Division of Engineering Services for developing the design and locations of the LAN closets.

Guidance and Support of Locally Managed LANs

NTG and PCB work closely to provide support for locally managed LANs. NTG engineers focus on the design and installation of the cabling system and appropriate hardware for setting up a network. PCB communications specialists provide guidance in the selection of the network operating system and the configuration of workstations and servers.

Groups considering the use of LAN technology are encouraged by DCRT to plan early for support staff to complement the support provided by DCRT. A LAN is a production service that cannot be effectively supported on an *ad hoc* basis. We recommend that a local server administrator be designated for each server or group of servers.

All Institute representatives are given a document developed by NTG entitled "Design Specifications for Local Area Network Communications Closets and Universal Building Wiring." This document provides a guideline for the implementation of wiring schemes within an organization on the NIH campus. It is intended to provide the user with a practical solution for minimum closet requirements and a structured building wiring system. Recommendations are given for LAN communication closets and horizontal and vertical wiring.

NTG is working closely with DES to include in the design of new and renovated buildings on campus provisions for networking. In Building 49, LAN closets, cable trays, and riser conduits have been incorporated into the building design. Our goal is to make the building architects see LANs as utilities and part of the infrastructure of the building. In addition, NTG is working with the NIH Telecommunications Branch on the wiring infrastructure for Building 49. Twisted-pair and fiber connections will be installed for voice and data applications.

Awards

This year the efforts of NTG were recognized with group and individual awards. David Songco, Harold Ostrow, and Joseph Kabara were recipients of a group award for the design and implementation of a sophisticated computer network system (RESnet) in support of laboratories, clinics and offices throughout NIH. This award was shared with Robert Brunelle, Roger Fajman, Frances Halverson, Robert Klein, and Oliver Morton of the Computer Center Branch, who collaborated on several aspects of NIHnet. Merike Kaeo and Jim Brunetti were recipients of Special Act awards for their individual efforts in the FDDI and CCnet implementations, respectively.

Where Do We Go Next?

We have identified a number of issues that should be addressed by DCRT and NIH management.

- Expansion: How do we add to our network and network services to meet increased demand?
- New technology: How do we track the new technology, and what technology is best for NIH?
- Support: How do we ramp up to meet the challenge of supporting an increasing number of users in a rapidly changing technology?
- Organization: How do we organize to accomplish all of the above?
- Funding and budget: How much will it cost to provide comprehensive networking services for NIH? How should users pay for network services? What are the options?

- Trans-NIH issues: What are the issues that cut across the NIH organizational structure? How should we fund network development and services? What is the role of the Network Policy Board? How can NIH organizations be involved in network policy and technical decisions?

In 1991 the activities of NTG were reviewed by the new DCRT Director. A formal peer review with experts from outside NIH was held at the end of June. This group reviewed the entire networking activities of DCRT and made recommendations for the best use of DCRT's networking expertise and resources. In 1992, the DCRT Director will implement his plan for the best mix of networking technologies for NIH and the management, operation and support of NIHnet. Work is now under way toward the creation of a new Computer Networking Branch within DCRT.

Projects

RESnet Backbone Technologies

M. Kaeo
with H. Ostrow, J. Young (DCRT/NTG)

During the last few years, high-speed communication "connectivity" between computers has become an increasingly important scientific resource on the NIH campus. The goal of the high-speed scientific research network, RESnet, is to tie together scientific research sites across NIH. In the past year, RESnet has evolved from a single Ethernet-over-fiber implementation to a backbone network comprising two separate Ethernet-over-fiber technologies and FDDI. A gradual phasing out of the proprietary 10 million bits per second Ethernet technology will take place as the 100 million bits per second FDDI network demonstrates its ongoing reliability.

The FDDI standard was one of the first standards to address LAN communications via fiber-optic media. Most of its components have been approved for over 2 years, and the standard is expected to be completed in its entirety by the end of 1991. In the past year, FDDI equipment has become widely available. The move to FDDI is occurring faster than originally anticipated because of the availability of FDDI equipment and NTG's goal of implementing standards.

In the past year, extensive work has been performed with the FDDI technology to gain the knowledge needed to support an FDDI campus backbone. Evaluation of the Cisco FDDI implementation was begun in September 1990. This consisted of testing the equipment's compliance to the FDDI standard, the capability of routing DCRT-supported protocols (DECnet, XNS and TCP/IP) over FDDI, and the capability of using both multimode and single-mode fiber-optic cable as the transmission media.

The move to the standardized FDDI technology brings the benefits of higher speed communication to workstations and hosts on individual LANs. In addition to the RESnet FDDI implementation, the expertise to recommend FDDI LAN solutions for scientists when they require networks running at FDDI speeds is available. Currently, two known applications for FDDI LANs on the NIH campus exist. One is the Advanced Laboratory Workstation (ALW) project. The second is the Magnetic Resonance Imaging Center, which involves the Diagnostic Radiology Research Program. Comprehensive interoperability tests will be performed with any new FDDI equipment before a connection to RESnet is made.

Presently, RESnet uses its own conduit system to provide the high-speed campus backbone to nine buildings on the NIH campus. Multimode fiber-optic cable, with 28 strands of fiber, is installed between each of these buildings. Expansion of RESnet to addi-

tional buildings will occur as new high-speed applications are identified. As new higher speed standards evolve, their progression will be closely tracked to keep up with changing technology and to provide the NIH community the network speed and service it requires.

Management of the RESnet Backbone

J. Young
with H. Ostrow, J. Kabara (DCRT/NTG)

In this ongoing project, NTG has utilized existing vendor and public-domain software on Unix workstations and is well on the way to implementing a comprehensive management strategy for the data networks at NIH. The strategy relies on the Simple Network Management Protocol (SNMP) which is part of the TCP/IP protocol suite. Because many vendors have accepted and implemented SNMP, their equipment can be monitored by software that makes use of the SNMP protocol.

Unix Workstations

Much of the software that is being developed to manage networks is being developed for Unix platforms. Each member of NTG has a Unix workstation that is part of the Advanced Laboratory Workstation project. The workstations are capable of running powerful network management software. Because the workstations are part of the ALW system, members of NTG have been able to share new software among the group on shared volumes, and have taken advantage of the backup facilities that the ALW project provides.

Network Management

NTG network management strategy is being implemented in several stages. The first stage of the project was to put a system in place to manage the networking equipment making up the different backbones of NIHnet. In September 1990, we evaluated a copy of the Cisco NetCentral network management package. This software has proven to be useful in the everyday management of RESnet.

Our NetCentral map includes objects from RESnet and from the other two large networks on campus (NUnet and NIAIDnet). NetCentral also monitors our connections to the Internet. The program is available to all members of the staff from their desktops. The package includes a net management screen which displays user-created icons to represent the state of individual network entities. The software polls each entity for its state at a user-defined interval. Through this screen, the user can add or delete entities to be queried by the software or query an individual entity for a range of values identified as the Management Information Base or MIB.

Another major feature of the NetCentral program is its database of information. The program retrieves and stores MIB variables from any number of the managed entities. The information is gathered and stored by a single workstation which has been specially configured as a database server. Information may then be retrieved and graphed by any station running the NetCentral package. The database also stores information such as serial numbers, warranty dates, and configurations for each individual network entity.

Managing Other Objects

The second stage of this project calls for managing objects which are not on the NIHnet backbones. The first stage of the project was straightforward, for many vendors had already implemented SNMP agent software in their devices. In the second stage of the project, NTG will compile agent software for key work-

stations and user services that make use of the NIHnet backbones. We will also try to discover other SNMP devices already on local networks. Key workstations might include the Silicon Graphics Workstations on the NIH campus, while key services might include the Convex minisupercomputer or the Domain Name Servers.

Future

Finally, NTG will need to explore ways to optimize its management strategy. Optimization may include removing SNMP traffic from the backbone and placing it on a dedicated network (out-of-band management). It may also involve running proxy agent software, which limits traffic by receiving all SNMP queries from the management stations and generating a response in proxy.

NTG plans to move to OSI when feasible. We need to explore a management strategy for the OSI protocol.

CCnet: Networking the Clinical Center

J. Brunetti
with H. Ostrow (DCRT/NTG)

The Warren G. Magnuson Clinical Center (Building 10), with over 2 million square feet and 5,600 people, presents unique networking problems. Since it is so large, the Clinical Center network can be considered as a separate campus backbone. We have designated this separate backbone "CCnet" for Clinical Center network. Eventually, CCnet will be a comprehensive cabling infrastructure servicing the entire Clinical Center with a mixture of fiber-optic and unshielded-twisted-pair cable. An interim infrastructure currently exists in Building 10. In February 1990, fiber was pulled up the west side MIS closet to every floor in the Clinical Center. A total of 13 28-strand 62.5/125-micron fiber-optic cables were run from room B1N239 to room N255 on each floor of the Clinical Center for a total of 364 vertical strands. Of this number, 234 have been terminated and tested (18 per floor). The remainder will be terminated on an as-needed basis.

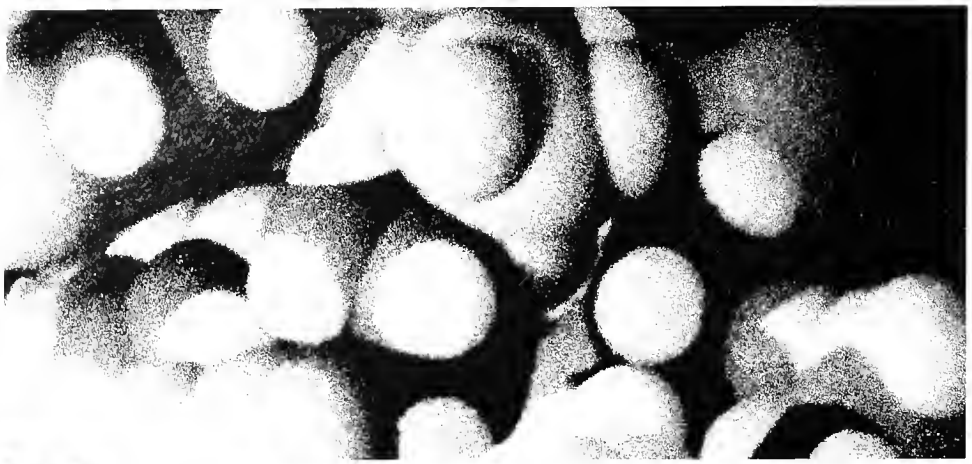
CCnet connects a total of 30 Ethernet networks from 12 different institutes, dispersed on every floor of the Clinical Center, to the NIHnet backbone. These networks are run over a variety of media. We have successfully connected Ethernets run over thick and thin coax, twisted-pair, and fiber cable to the backbone.

There are two special-purpose all-fiber networks connected to CCnet. The first connects three RITA x-ray digitizers on different floors. These machines digitize and transmit x-ray images. This network is totally internal to the Clinical Center, and no outside connection exists. This is the start of a comprehensive imaging network for the Clinical Center.

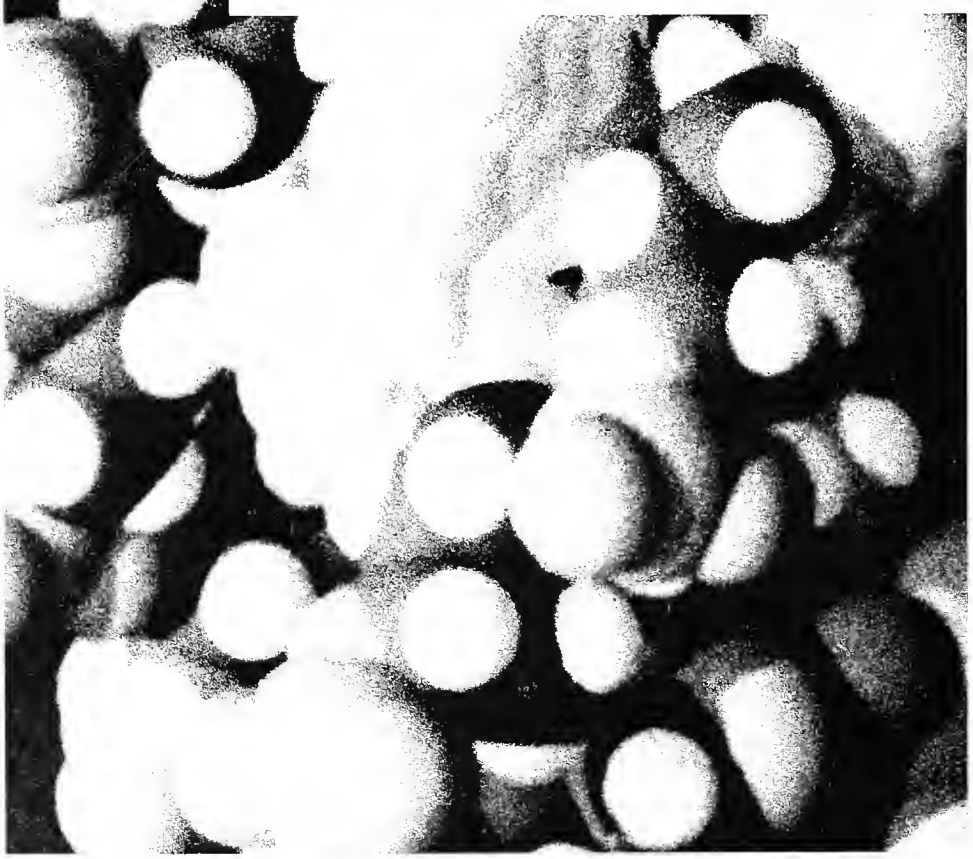
The second all-fiber network is called the "Orphan Network." This network was originally installed to connect five Silicon Graphics machines to the campus network. These five machines belonged to scientists who had no departmental LANs to connect to, but desperately needed connectivity to the Convex. The Orphan Network has grown to 20 machines belonging to six different institutes on nine different floors. Our policy for connecting orphans is to connect only one machine per branch. This is the carrot to entice the branch to put in a comprehensive network. We strongly discourage connecting any more machines from that branch until a LAN has been installed. When the network is in place, the fiber used to connect the orphan is used to connect the new network to its own router connection.

The existing infrastructure in the Clinical Center

is only an interim solution. A more permanent infrastructure is planned. Three LAN closets each will be built on floors 1-13 for a total of 39 closets. Space for these closets has already been identified and allocated.



PCB



Personal Computing Branch

David C. Songco, Chief

Executive Summary

Fiscal Year 1991 was another year of precipitous change in the personal computer industry and a particularly challenging year for the Personal Computing Branch (PCB) in its ongoing effort to support the growing legions of microcomputer users at NIH.

Macintosh usage at NIH, though still far below that of Intel-based PCs, continued to grow as more users opted for its relatively easy-to-use graphical interface. Among the Macintosh products PCB evaluated and announced support for were the new System 7 operating system, multimedia authoring tools, network-based programs, AU/X, and a variety of hardware devices. On the PC side, PCB announced support for a wide range of new desktop and portable systems as well as a number of upgrade options for older systems. PCB also began actively promoting Microsoft Windows, which brings a graphical user interface to DOS. We also recommended users upgrade to DOS 5.0, which was released in June.

Change was especially fast-paced in the network arena, where PCB continued its evaluation of next-generation network operating systems and mail packages. One result was our recommendation of Microsoft LAN Manager for new networks. In addition, PCB did its best to promote AppleTalk support on the campus-wide network. Evaluation of client-server database products also continued.

Training—another mainstay of PCB support—continued in FY91 at roughly the same level as in FY90. Approximately one-sixth the number of NIH personnel attended one of nearly 250 course sessions offered. Meanwhile, the NIH User Resource Center, a walk-in computer facility cosponsored by PCB and the NIH Training Center, experienced a 50-percent increase in clients in FY91.

A notable new initiative in FY91 was increased support for scientific computing, made possible in large part by the hiring of new PCB staff with scientific backgrounds. We plan to further strengthen our support for scientific computing in the coming year.

The Personal Computing Branch has a staff of nearly two dozen highly trained computer specialists whose mission is the ongoing support of PC and Macintosh computer users at the National Institutes of Health. PCB support encompasses the evaluation and recommendation of select hardware and software products; training in their use; and problem resolution through the PCB Help Desk and related services.

Anyone who has ever had to buy computer hardware or software knows there is a bewildering variety of products on the market today, and it is not always easy deciding which product is best. To help NIH computer users make the most of current computer technology, PCB evaluates promising new products in many areas and comes up with recommendations, which we share with the NIH community through a number of vehicles.

The NIH computer user's best guide to computer hardware and software is the PCB Product Information Guide. Available in separate PC and Macintosh versions, it exhaustively describes all PCB-supported products and offers usage suggestions. The Product Information Guide is updated regularly. NIH computer users can get the latest copy of either document by calling the PCB or by downloading it from the PCB's electronic bulletin board service, PCBull.

PCB regularly publishes and distributes PCBriefs, a 12- to 16-page technical bulletin for NIH computer users. Each issue contains feature articles, product updates and informative reviews, training news and schedules, and a helpful question-and-answer section. PCB also disseminates computer-related information through campus user groups it sponsors or participates

in: the Biomedical Research Macintosh Users Group, 3Com Campus Users Research Exchange, and the NIH Campus WordPerfect Working Group.

To relay information to NIH users as quickly as possible, PCB maintains a PC-based electronic bulletin board system called PCBull. Besides containing a wealth of timely computer-related information, PCBull has hundreds of useful PC and Macintosh software programs free for the downloading. It is also host to a dozen forums for discussion of computer-related topics, where users can post queries for PCB staff and other users. All one needs to use PCBull is a modem and communications program.

Personal computers are getting easier to use, but most people still need training to use them effectively. DCRT, in conjunction with the NIH Training Center, teaches a full spectrum of computer courses, ranging from introductions to DOS, Windows, and Macintosh, to courses dealing with specific applications. More than 2,000 persons take DCRT-sponsored training each year.

Wherever personal computer users are at NIH, as long as they have a telephone, they have access to first-class computer support. Users having a problem with any PCB-supported product can call the PCB Help Desk to talk with a PCB support specialist. For onsite support, they can see their local ICD-appointed, PCB-trained Lead User or Macintosh Support Coordinator.

Technology Tracking

Personal computer users at NIH look to the PCB for guidance in the selection of computer hardware and software. Keeping up with the latest developments and predicting future trends is becoming an increasingly challenging task, particularly with respect to PC and local area network technologies, which are currently in a state of constant flux.

Macintosh

In FY91, two new moderately priced Macintosh models of particular interest to NIH users were introduced, the LC and IIsi. Lower prices combined with increased recognition of the Macintosh as a powerful yet easy-to-use computer resulted in yet another year of significant growth in Macintosh usage at NIH. Major software vendors such as Lotus Development Corp. and WordPerfect Corp. announced releases of their software for the Macintosh, further increasing the Mac's attractiveness in a mixed PC/Mac environment.

System 7, the long-awaited Macintosh operating system upgrade, was released in FY91. Although PCB began supporting System 7, we held off recommending users upgrade mainly because of its large memory requirements (4MB) and the fact that many current applications need to be upgraded to be System 7 compatible. Fiscal Year 1991 was a year of planning and preparation for Mac users at NIH. Besides disseminating information on System 7 through NIH-wide seminars, presentations, and articles, PCB encouraged users to prepare for the upgrade by purchasing the required memory and by identifying software that will need to be upgraded. In FY92, PCB will actively encourage Mac users to switch to System 7. By then, training and applications that take advantage of System 7 ought to be available.

The Macintosh's reputation as a powerful multimedia tool was enhanced in FY91 with the release of new, easier-to-use packages such as MediaTracks. PCB continues to track developments in multimedia technology and to consult in the use of these packages.

PCB also continued to track the profusion of third-party hardware devices for Macintosh computers. New font technology, TrueType, became available in FY91. TrueType enables the effective use of a new crop of moderately priced yet high-quality non-

PostScript laser printers. Growing out of FY91 product evaluations were new PCB recommendations for high-capacity digital audio tape drives for backup, low-radiation monitors, removable cartridge drives, scanners, printers, and portable hard disks.

PCB also evaluated both TCP (transmission control protocol) and AppleTalk network-based programs for the Macintosh, and began accepting consultations on applications in both areas. TCP programs evaluated include products for remote login, FTP (file transfer protocol) file transfer, as well as various Internet access tools. AppleTalk programs included AppleShare, System 7 file sharing, and network-based "groupware" applications such as MeetingMaker, a scheduling program that allows work groups to electronically coordinate meeting schedules, conference room bookings, and personal calendars.

In FY91, PCB increased its involvement with the Unix operating system by establishing a dedicated AU/X (Apple's version of Unix) Macintosh, which it used as a test site for FTP distribution of Macintosh public-domain software to selected users and as a server for database testing. While we consider AU/X a reasonable choice for a low-cost Unix workstation where performance is not critical, its graphical interface is too slow on Macintoshes other than the high-end IIfx. For this reason, PCB began investigating the integration of high-performance Unix-based Sun workstation technology into the Macintosh environment. Working in collaboration with other DCRT groups, we evaluated Unix-based backup systems, file servers, database servers, and Macintosh front-end programs. We plan to continue these efforts in FY92.

PC

On the PC side too, advances in technology brought increased performance and greater capabilities at a lower cost last year. During FY91, PCB evaluated and announced support for what it considers an entry-level system based on a 386SX processor; new 386-based models, which we now recommend for most applications; and new high-end 486-based models, which due to their relative high price are still of limited appeal to most NIH users. Because of declining prices of 386-based PCs and the need for 386 or better performance to effectively use state-of-the-art graphical user interfaces for the PC such as Microsoft Windows, we no longer recommend the purchase of 286-based machines. For users of 286-based systems who are unable to buy replacement systems, PCB explored and announced support for a range of upgrade options, such as Intel's SnapIn 386 processor replacement module, IBM's high-performance, high-capacity SCSI-based disk subsystem, and high-resolution video systems, including ATI Technology's 8514/Ultra adapter and NEC's MultiSync 3D and 4D displays.

Interest in portable computers continues to rise as desktop computer capabilities are bundled into increasingly smaller packages. To the already extensive PCB lineup of supported models, ranging from the tiny Compaq LTE through the full-featured PS/2 Model P70, we added two new laptops in FY91: the Compaq LTE 386s/20 and IBM L40 SX, both of which offer among the best balance of power and portability in the industry.

In response to growing requests for more low-cost systems, PCB began an intensive evaluation of PC compatibles "clones." In deciding which makes to evaluate, we are taking into consideration such factors

as price, diversity of product line, vendor support, upgrade potential, and standings in PC industry ratings. The systems chosen for hands-on evaluation will be subjected to standard benchmarks and thoroughly tested with PCB-recommended third-party hardware and software, including Windows and OS/2. We expect to have preliminary results by early FY92.

PCB evaluation of the two leading PC graphical user interfaces (GUIs)—Windows 3.0 and OS/2—continued in FY91. Although the advanced features of OS/2 continued to hold much promise, the lack of major OS/2 applications and uncertainty about OS/2's future led PCB to recommend that users wishing to move to a graphical PC interface adopt Windows, which appears viable and now enjoys a full suite of applications, including Windows versions of Lotus 1-2-3 and WordPerfect. PCB now believes that easy-to-use GUIs are the future of PC computing, and although we currently recommend Windows over OS/2 for most users, we will continue to track OS/2, a new version of which is scheduled to be released early in FY92.

Meanwhile, DOS users were treated to a major new release in FY91: DOS 5.0, featuring advanced memory-management capabilities, a GUI-like shell, online help, and many new commands and options. PCB announced support for both PC and MS-DOS versions and recommended most users upgrade.

Networking

During FY91, PCB continued its evaluation of next-generation network operating systems with the potential for eventually replacing 3Com 3+Share at NIH, including the OS/2-based LAN Manager from Microsoft, Novell's NetWare386, and AppleShare for the Macintosh. Due to the uncertainty in the industry, our recommendation to current 3+Share users was to stay with 3+Share if possible. To those planning to install new networks, we recommended LAN Manager. PCB also evaluated several next-generation electronic mail systems, including Lotus cc:Mail.

Currently the AppleTalk communications protocol—built in to every Macintosh—is not supported on the burgeoning NIHnet. In FY91, PCB championed the effort to evaluate and test AppleTalk by establishing, then chairing, a committee of DCRT technical communication experts from CCB, CSL, NTG, and PCB. The evaluation concluded with an agreement in principle to support AppleTalk and to proceed with a staged implementation test of AppleTalk backbone support while concurrently working with DCRT management to find the resources necessary to support AppleTalk on a wider basis. Indeed the future of Macintosh computing at NIH is tightly interwoven with the ability of distributed Macintoshes to communicate and share resources with one another using this protocol.

Database Management

PCB continued to provide consulting support for users of DBASE IR Plus and stepped up its evaluation of other database management packages that could function well in a client-server environment, where networked "front-end" applications, or clients, access database applications, or "back-ends," running on servers. Client products evaluated thus far include Borland's Paradox and Paradox SQL Link, as well as products from Oracle Corp. Server products evaluated include Microsoft's SQL Server, and Oracle's OS/2 Server and products that allow access to a DB2 mainframe server. We expect client-server technology to play a strategic role in promoting information interchange at NIH.

In a similar vein, PCB participated in the evalua-

tion of Oracle as a possible solution to the problem of moving the extramural grant and research contract information database (Information for Management Planning, Analysis and Coordination—IMPAC) to a relational database management system. Part of the evaluation process involved working collaboratively with CCB and DMB in evaluating Oracle products running on micro-based clients and servers, as well as Oracle's access to DB2 on the MVS-based 3090 mainframe.

Training

Besides offering guidance in the selection of computer products, PCB, in collaboration with the NIH Training Center and the DCRT Training Unit, continued its efforts to provide users with high-quality training in the use of supported products. Training figures for FY91 were nearly identical to those for FY90: A total of 131 PC-based and 112 Macintosh-based personal computing course sessions were presented under the auspices of the NIH Training Center. Of those, PCB cosponsored 155 sessions of 12 courses, which were attended by 2,390 students. Additionally, 16 sessions of 3 courses (with 256 students) and 16 limited-focus seminars were presented through the DCRT Training Unit without fee. In all, PCB staff taught 24 percent of all courses and seminars offered, while outside vendors, under the supervision of the PCB and using PCB-developed materials, taught the rest.

In FY91 major revisions were made to nine PCB-sponsored courses, resulting in new course manuals and course scheduling. Notably, 3Com network training was revamped and expanded to provide an advanced user course as well as a network management course. Both courses were firsts of their kind.

As in previous years, the PCB Associate Instructor Program played a large role in the success of our training efforts. Under the program, experienced NIH computer users volunteer their time assisting instructors during hands-on training courses. During FY91, 89 persons from 28 ICDs participated in 155 course sessions.

In addition, the NIH User Resource Center (URC) continued to serve as a vital adjunct to PCB training services. Sponsored by PCB in collaboration with the NIH Training Center, the URC is a multipurpose walk-in computer facility equipped with PC and Macintosh workstations, an extensive collection of applications software, and a variety of peripherals such as laser printers, modems, CD-ROM players, and page scanners. It also has a large selection of self-study courses and many popular personal computing periodicals, books, catalogs, and other publications. During FY91, NIH employees made more than 6,000 visits to the URC—up from 4,000 in FY90—for such purposes as researching computer topics, evaluating PCB-supported hardware and software, taking self-study courses, and consulting with URC staff. The URC Teaming Assignment Program, under which NIH employees volunteer 4 hours of their time per week for 3 months in exchange for the opportunity to enhance their computer skills by working directly with URC staff, remained popular.

Consulting Services

To help computer users when they encounter problems, PCB maintains a telephone help desk. When users call, a dispatcher enters their name and a description of the problem into a database system, which in turn generates an electronic mail message to the appropriate PCB specialist, whose responsibility it is to returned the user's call within 24 hours. Most calls are return within 30 minutes. Additional consulting requests come in through the PCB's electronic

bulletin board and directly to PCB staff via electronic mail from certain designated user groups. During FY91 we responded to more than 6,000 requests for assistance.

To further enhance their Macintosh consulting skills, some PCB staff in FY91 took special Apple-sponsored training to become Apple-certified repair technicians. This expertise proved to be helpful in diagnosing Macintosh hardware problems. PCB does not offer repair services to NIH but is working with other NIH groups who may provide those services in the future. We are also working on oversight programs to help monitor and improve the often uneven repair service provided by outside vendors.

In an ongoing effort to improve our consulting services, PCB conducted a telephone survey of our user community in FY91 and came up with several recommendations: to respond more quickly to user questions, perhaps by staffing the help desk with specialists; to accept more requests for help via electronic mail; to involve the User Resource Center more heavily in the consulting process; and to increase the training of PCB Lead Users and Macintosh Support Coordinators.

Increasingly in FY91 PCB relied on its Lead Users and Macintosh Support Coordinators to handle routine problems at the local ICD level. In exchange for the support they provided on behalf of PCB, Lead Users and Macintosh Support Coordinators, which numbered about 200 and 90, respectively, received free training and priority access to PCB staff. PCB also held regular meetings for these groups, at which topics of particular interest were presented, and we sponsored special Macintosh training classes at Apple's facilities in Reston. Also in FY91, 20 Lead Users were designated Certified Lead Users in recognition of their special contributions. PCB continued its effort to make ICD management aware of the vital function these persons perform.

Ongoing PCB sponsorship of a number of computer user groups also contributed significantly to its support mission in FY91. Meeting monthly since January 1990, the Campus Users Research Exchange (The CURE) brings together network administrators and users from across campus for network-related presentations and information exchange. Another network group, the Server Administrators (SADs), which consists of all DCRT 3Com server administrators, meets twice a month to discuss not only operations of the DCRT 3Com network, but also future directions for DCRT LAN support at NIH. Monthly meetings of the PCB-led Biomedical Research Macintosh Users Group (BRMUG) continued to draw large audiences with timely presentations on such topics as System 7 and techniques for scientific poster-making. Meanwhile, PCB turned over leadership of the popular NIH WordPerfect Working Group to Lead Users from NHLBI, NLM, and OD, although it continues to participate.

Increasingly, PCB responded to requests for in-depth consultation in FY91. For example, it worked collaboratively with DCRT's Network Task Group on several occasions to provide NIH groups with detailed recommendations on local network design, cabling requirements, network operating systems, servers, and applications, thus ensuring that the systems they implement today will be able to participate in future campuswide connectivity schemes. PCB also began an ongoing collaboration with NCI's Clinical Oncology Branch in the design of a comprehensive Macintosh-based client-server database system. We expect to do more such organizational consultations in FY92.

Information Dissemination

In addition to relaying timely computer-related information to key NIH persons through Lead Users, Macintosh Support Coordinators, and user groups, PCB continued its formal vehicles for NIH-wide distribution of computer-related information: PCBriefs, PCBull, and the PCB Product Information Guide.

In FY91, five issues of PCBriefs, the PCB's technical newsletter for NIH microcomputer users, were published and distributed to 7,000 NIHer's. A special additional mailing of the scientific application-oriented May 1991 issue and the August 1991 issue went to 1,500 NIH scientists to increase their awareness of PCB-supported products and services. Also in FY91, we published two editions each of the PC and Macintosh versions of the PCB Product Information Guide, which lists and describes all PCB-supported hardware and software products. PCB also prepared and distributed an instructional booklet showing researchers how to use campus mainframes for downloading GenBank sequences (Downloading GenBank Sequences for Use with Sequence Analysis Programs) and presented a course on the subject.

PCBull, the PCB's electronic bulletin board service, remained popular in FY91, averaging 1,100 calls per month—about the same number as in FY90. Accessible 24 hours a day from personal workstations via telephone communications, PCBull is an integral part of the PCB support program, offering NIH users information, software updates, and tips on using PCB-supported products, as well as PCB utilities, publications, and useful public-domain files for the PC and Macintosh. In addition, users can ask questions online as an adjunct to the PCB help line. A new GUI (graphical user interface) Conference dealing with Windows and OS/2 was added to PCBULL during FY91, as was the SCIENCE Conference, which is devoted to discussion of scientific PC hardware and software.

Increasingly in FY91, information was distributed to NIH computer users through electronic mail groups. As LAN connections proliferate, we expect this channel to become even more useful. As always, the User Resource Center played a large part in the dissemination of PCB information.

Other Support Services

PCB continues to maintain the local DCRT 3Com network, induding the DCRT network in Building 31. This means managing its daily operations as well as its internetwork connections to more than 100 network domains. Maintenance involves working with the CCB mail hub, the Network Task Group, the server administrators from each of the labs and branches of DCRT, the network administrators from the NIH ICDs, and other federal agency network administrators.

PUBnet, Fax Hub

During FY91, PCB worked to further develop the Public Network (PUBnet), a stand-alone network designed to provide various network services to users of NIH LANs connected to the NIHnet (RESnet, NUnet, CCnet, etc.). PUBnet connects to the NIHnet via RESnet, and also has limited asynchronous dial-in capabilities to service those persons not yet attached to a LAN or NIHnet. It consists of a number of network servers that support both PCs and Macintoshes. Users can link up to these servers from their own workstations and access, download, or run tasks directly from PUBnet. A menu interface guides users through the available services. Users familiar with disk structures may choose to bypass the menus

and explore the subdirectories directly. Especially popular in FY91 was PUBnet's fax gateway, which acts as a hub over the NIHnet, giving 3Com LAN users at NIH the ability to send electronic faxes from their own network mail to any fax machine nationally or internationally.

Also in FY91, the DCRT Library moved all its electronic information over to PUBnet, including library news, acquisitions lists, online catalog information, and CD-ROM subscriptions. Future plans for PUBnet include making more ICD-supplied documents available. In addition, Personnet, the personnel system regulations on CD-ROM, will become available early in FY92. With additional gigabytes of disk space, PUBnet will become a major repository of NIH-wide information for high-speed distribution across campus.

Scientific Computing Support

In FY91 PCB also undertook a major new initiative to strengthen its support for scientific computing at NIH. Led by two new PCB staff members with backgrounds in molecular biology, this took the form of numerous product evaluations and announced support for specific scientific applications, such as bibliographic managers and programs for doing DNA and protein sequence analysis. Together with NCRR's Medical Arts and Photography Branch, PCB published a booklet detailing the means of producing multiple- and single-piece scientific posters on a Macintosh at greatly reduced costs compared with traditional means; PCB staff themselves used those techniques to prepare several posters for NIH Research Festival '91 in September. PCB staff worked closely with DCRT scientists to develop plans for an NIH Scientific Computing Resource Center.

Computer Security

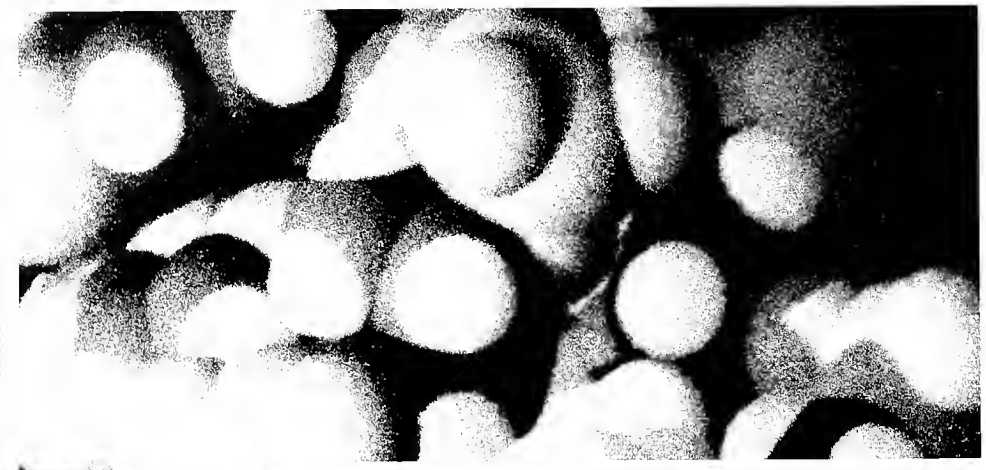
In the face of a growing and potentially serious threat posed by computer viruses, PCB redoubled its effort to increase the security awareness of personal computer users on the NIH campus. During FY91, PCB's Security Coordinator initiated efforts to instruct personal computer users in safe computing practices and promoted the use of PCB-supported antiviral software. Instructional documents and informational bulletins were updated to reflect new viruses appearing on campus, and support was announced for new antiviral software and new versions of previously supported products. A number of lectures and classes on the subject were also offered. Finally, the PCB began evaluation of other security products, such as data encryption and password-protection programs on both Macintosh and PC platforms.

Future Directions

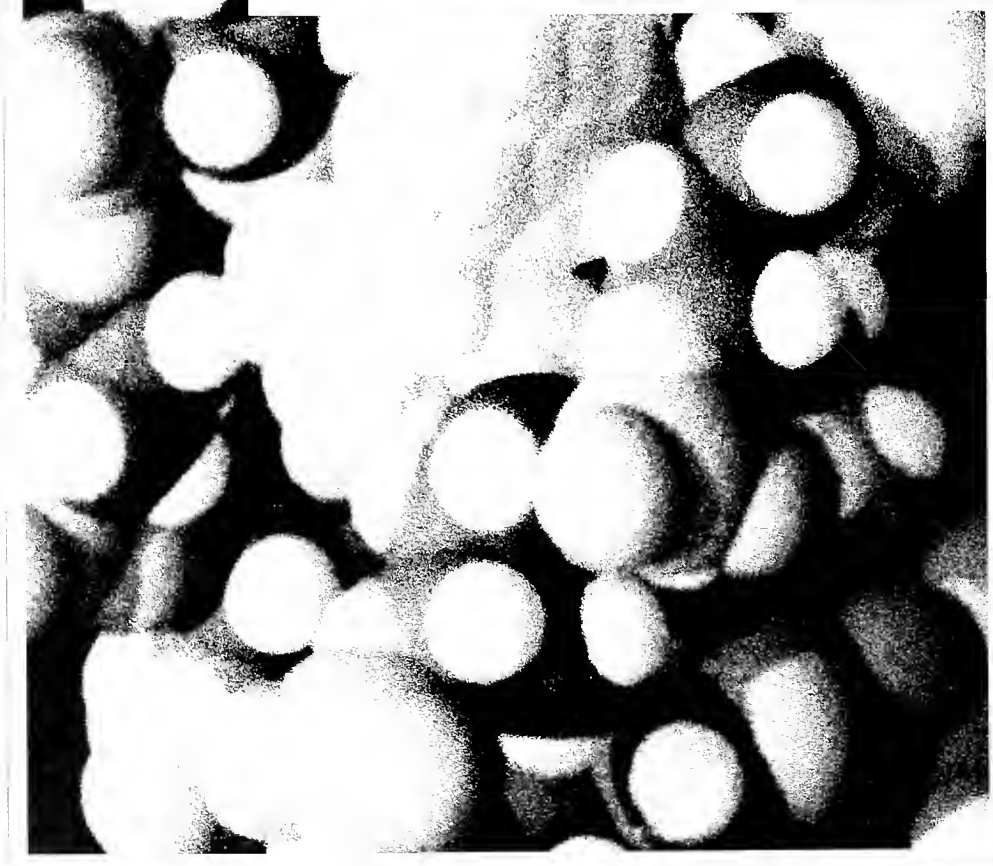
FY91 marked the 10th anniversary of personal computing as we know it. Each new year brings exciting advances in PC technology, and it seems the pace of technological change is accelerating. Keeping on top of the PC industry these days is no mean feat.

FY92 is likely to be a critical year for PC operating systems, answering the question once and for all: Does the PC future belong to Windows or OS/2? On the Macintosh, System 7 will almost certainly establish itself. We also hope to see clearer direction in the area of network operating systems. Client-server and distributed computing technologies can be expected to continue their steady evolution, although it will clearly be several more years before they come of age. PCB, in collaboration with other DCRT groups, will significantly strengthen its support for scientific com-

puting with some bold new initiatives in FY92, including launching of a new NIH Scientific Computing Resource Center. Despite the challenges we face in dealing with this fast-paced industry, it is with great enthusiasm that PCB welcomes the second decade of the personal computer.



DMB



Data Management Branch

J. Emmett Ward, Chief

Executive Summary

The Branch implemented personal property accountability, travel voucher processing, online funds formulation and funds certification and the government-wide standard general ledger as new systems under the umbrella of the Administrative Data Base. Expansion of the travel system to include foreign, local and reimbursable travel is planned for FY92. Reconciliation of property accounts and electronic processing of property passes are also planned for FY92. Investigation of new information processing, the use of SQL servers and data interpretation systems has been initiated in the recently formed New Technology Analysis Section.

A subcommittee of the ADB Steering Committee was established during FY92 with 70 representatives from the ICDs to develop an ADB information requirements document. This document will be presented to the Steering Committee in its October 1991 meeting. Plans call for a phased implementation of these requirements beginning in Spring 1992. Other significant computer applications supported by the Branch include the Clinical Information Utility, AIDS Research Information System, Child Health Information Portfolio, Blood Typing and Donor/Patient Antigen Matching, and AIDS Loan Repayment.

The Data Management Branch (DMB) provides advice and assistance to research investigators, program officials, and administrators throughout NIH in planning for and obtaining computer data processing services. A central NIH resource for systems analysis, design, and programming, the branch also develops, maintains, and processes the NIH Administrative Database and the Clinical Center's Clinical Information Utility. On the staff are 44 permanent full-time employees whose disciplines include computer science, mathematics, and statistics.

DMB staff members design and create computer-based data management systems that provide practical solutions to the unique mix of administrative, scientific, and management data processing problems encountered at NIH. Comprehensive training in the facilities and functions of all systems developed by DMB is provided to each new computer system user.

Three sections make up the DMB. The *Applied Systems Programming Section (ASPS)* provides general computer systems analysis and programming services to all of the ICDs. The *New Technology Analysis Section (NTAS)* was established during FY91. An overview of the purpose of the NTAS is presented in the next paragraph. The *Database Applications Section (DBAS)* develops and maintains the NIH Administrative Data Base, which provides broad support for all administrative and financial processes at the NIH.

As a result of the many technological changes in database and information management and of the evolving requirement for computational support across microcomputing and mainframe platforms, the DMB has shifted some of its direct service staff to the NTAS. Along with this shift, the services previously offered by the Scientific Applications Section (SAS) have been absorbed by the ASPS, and the SAS has been eliminated. The NTAS brings together branch personnel with expertise in hierarchical and relational database management concepts as well as many years of computational and collaborative experience with NIH administrators, managers and scientists. This new Section will be responsible for analyzing and selecting new database management approaches and for developing the methodologies which will facilitate their use across multiple platforms.

Broad Involvement in the NIH Mission

The NIH Administrative Data Base (ADB) represents a major effort by the NIH to integrate the administrative and financial data of its intramural program. A special approach to database management, the ADB concentrates on the full sharing of data among all subsystems that support the NIH intramural program. It features online point-of-origin data entry, minimized data redundancy, background generation of all accounting transactions, and fully synchronized information processing (i.e., what you see is the latest state of any process, data or function).

The ADB is an ongoing developmental project that encompasses the following features:

- it provides full support for the purchasing, receiving and payment of goods and services
- items in nine inventories are individually tracked and are made available by way of online stock requisitions and are completely integrated with the self service stores operation
- all vendors, vendor credits and vendor source agreements are maintained and tracked
- NIH cashier functions are fully supported
- some 13 service and supply fund activities have been or are being integrated
- domestic travel orders and travel vouchers are tracked and processed with local and foreign travel being developed
- the major development and transition to a new NIH-wide property management system has recently been completed
- the implementation of full research contracts support is under development and accounting functions such as funds formulation and funds certification have been shifted to online ADB support. A more specific summary of new ADB initiatives during FY91 follows.

The AIDS Loan Repayment System was developed for the Office of AIDS Research, OD/NIH to support the repayment of outstanding student loans for scientists who are conducting AIDS-related research at NIH. This system is integrated into the ADB and utilizes its procurement, invoice and accounts payable functions. Specifically, the system supports online data entry for new recipients, obligation of loan and interest balances, invoice preparation, payment approval, treasury check processing, payment history, federal tax payments and processing of IRS 1099 forms.

Phase I of the Property Management System enabled the conduct of a property inventory of all buildings on and off the NIH campus, and the reconciliation of the new and old property accountability systems. This phase is in the reconciliation stage. All inventoried items have been loaded into the new Property Database. New acquisitions are automatically entered into the Property Database by way of the ADB receiving module. Currently, the system will handle new acquisitions, property transfers and surplus, review of undecaled items and modifications of property items. During the coming year, support will be provided for online creation of property passes, personal custody forms and property loan forms.

The travel voucher process for domestic travel orders was completed during FY91. This system presents the ICD travel official with a proposed travel voucher based on the original travel order. The travel official is permitted to modify the voucher based on actual travel performance and expenses. The voucher is then electronically released to DFM for final review and payment. The domestic travel system will become mandatory during FY92.

Online formulation of appropriated funds is now available in the ADB. This facility enables the appor-

tionment of funds down to the allowance level. Online fund certification of committed and obligated amounts has also been added. During any ADB process which affects commitments and obligations, allowance balances are automatically verified. If a transaction causes an allowance balance to become negative, a "once-daily" warning message is automatically queued to the appropriate budget office. This process is completed in a transparent manner to the committing/obligating function and that process is not interrupted.

The most difficult conversion during FY91 has been the tedious task of moving to the new government-wide standard general ledger. This process involved not only the development of a new decode table for existing transactions, but also the conversion of many old processes to either new or different transactions. This required the reconciliation of old and new effects in many situations that did not have previously existing processes—a tedious effort. Final parallel testing and manual cross-footing appears to have resulted in a fully acceptable new standard general ledger. Plans call for FY91 reporting and FY92 processing under this new system.

In addition to the ADB development efforts of FY91, several major modifications were needed to implement legislated requirements and to improve access, security and accuracy of ADB data. During FY90, DMB brought contractor staff onboard and with their support, several major maintenance tasks were completed in a timely and accurate fashion. One of the maintenance tasks in the procurement area required that more than two dozen software modules and 28 screen formats be modified within a short timeframe.

Other DMB-Supported Projects

Developed during the 1970's as an historical archive of clinical information for research, the Clinical Informational Utility (CIU) gathers data from the Medical Information System (MIS), the Medical Records Department and the various service organizations in the Clinical Center (CC). Over the years, millions of records have been archived and made available for use in ongoing research protocols and for retrospective search and display. Over 95% of *ad hoc* requests are satisfied with overnight turnaround. The CC Information Systems Department monitors and authorizes all users of CIU data, and the CIU automatically tracks and reports each access of the database. To satisfy clinical investigator needs, the CIU currently handles approximately 10 recurring and 20 *ad hoc* requests each week. The CIU is maintained by personnel in the Office of the Chief, DMB. Current efforts are centered on integrating CIU data classes and permitting overnight retrieval among them.

While providing support for protein-DNA interaction research, the need became apparent for a powerful and user-friendly molecular dynamics system on campus. After serious investigation, DMB began evaluation of several BIOSYM modules. These modules were subsequently installed on the DCRT Silicon Graphics facility and testing was conducted using sample modeling problems. Two modules have been installed and are being supported. DMB is now maintaining the BIOSYM software and providing consultation regarding its use. Future plans call for including BIOSYM as part of DCRT's planned Scientific Computing Resource Center. Also during the year, DMB continued to provide support for sequence services by way of DNASTAR, PC/GENE, GCG and the

Intelligenetics Gene Bank On-Line Service (GOS) systems.

DMB continued to provide software support for the AIDS Research Information System of the Office of AIDS Research. This system reports research expenditures in the ICDs using the Mason research category codes and special interest codes. Quarterly reports of expenditures are generated, with details of this information stored in an online relational database to support *ad hoc* queries. Input of expenditure data is received by way of different media (PCs and mainframes) and different modes (mainframe files, SYMPHONY databases and LOTUS 1-2-3 databases). Data are converted to a standard mainframe format, summary reports are generated, and a relational database of the detail is created for online query using SQL. During FY91, the retrieval facility was completely redone to shield users from the need to use SQL. A comprehensive front end was developed with menu presentations and help facilities that guide the user through the maze of accessing and summarizing category and expenditure data.

The Child Health Information Portfolio System (CHIPS) provides a central facility for timely and easy access to IMPAC and NICHD-specific data for grants, pending applications, jointly funded awards, and subproject and intramural grants for current and all past fiscal years. Data retrieval and display criteria are dynamically defined by the users as the need arises. While CHIPS is available to all NICHD personnel and provides tools for the analysis and management of grants data, it is currently used most extensively to support the functions of the Office of Science Policy and Analysis/Program Analysis Branch, the Office of Grants and Contracts, and the extramural research staff. Future development of CHIPS includes:

- a system to assign and track minority supplements
- an applications tracking system servicing all ICDs
- production of council letters and labels
- analysis and development of mechanisms supporting Trainee Appointment Data, Payback Data, FHAs FTTPs, RFAs
- a procedure for the dynamic inclusion of additional NICHD-specific data items.

The Blood Typing and Donor/Patient Antigen Matching System was developed for the Laboratory Service Section of the Department of Transfusion Medicine, NIH Blood Bank. The system provides (1) enter, edit and update of donor/patient antigen-typed blood to an online database (2) capabilities for listing, query and transfer of inactive donors/patients and (3) an archival database for inactive donors/patients. At the present time, a DB2 system is being developed for normal volunteer data. Interactive System Productivity Facility (ISPF), DBEDIT and Query Management Facility (QMF) procedures are being written to provide a system that has full-screen data entry with online editing and standard reports. The systems' functions will be invoked via an ISPF panel.

DMB continues to provide leadership, consulting and technical support on the development of new and improved stocking criteria for the NIH Central Stores inventory. Critical parameters such as average sales, demand history, rate of increase, number of customers, predictable demand, cost data, overhead cost, and storage cost are monitored with complex statistical methodologies. The quality control for these parameters provides improvement in reliability of cost avoidance measurements.

To support Veteran's Administration Medical Centers, DMB has completed work on complex algorithms that compute trends in decubitus ulcer rates and compare them with peer group hospitals and national rates. We have also prepared a set of graphs for each of the Veteran's Administration medical centers displaying the comparative results.

To support the efforts of NIMH's Dr. Norman Rosenthal in a study to distinguish summer depressives from winter depressives, DMB developed a database of Hamilton Rating Scale (HRS) scores and Seasonal Screening Questionnaire (SSQ) scores. Most of the analysis involved stepwise discriminant analysis. These programs were highly successful in identifying and distinguishing the two groups of patients.

To support interview scheduling of applicants to the Commissioned Corps of the PHS, major modifications were made to the WYLBUR command procedures. New data entry formats were developed and the scheduling programs were modified. This process is now complete and the system is being productively used by clinical investigators.

Evaluation of Database Management Technologies

During FY91, DMB staff collaborated with staff of the Personal Computing Branch (PCB) and the Computer Center Branch (CCB) in the analysis of ORACLE software with the DB2 mainframe relational database management system. NTAS staff installed the ORACLE personal computing products and tested them with the ORACLE Relational Database Management Systems (RDBMS) on the LAN server and with DB2 on the mainframe platform. NTAS staff also developed the test applications for the local and server RDBMS evaluation.

Following a detailed evaluation involving outside collaborations and investigations as well as collaboration with key CCB staff, DMB determined that a full conversion of ADB processing from hierarchical to relational database technology was not recommended at this time. However, an effort to convert appropriate ADB data to relational format for information processing was recommended. With this in mind, a subcommittee of the ADB Steering Committee has been formed. This subcommittee comprises members from administrative, budget/financial, laboratories/branches, procurement and technical staffs of the ICDs. The subcommittee is charged with the responsibility of identifying information requirements from the ADB and of establishing the requirements for information access and display. The subcommittee is to make its final report to the ADB Steering Committee in late October 1991.

An evaluation of 66 personal computer relational database management systems (RDBMSs) was completed during the year. From this group, four RDBMSs have been selected for full evaluation. The procurement process will be completed by the end of FY91 and at the same time, an SQL server facility will be installed on the DMB OS2 server. A full evaluation of these RDBMSs will begin early in FY92.

As an interim solution for those who have mainframe relational databases and require a graphical user interface for data retrieval as well as a full set of personal computing facilities on the LAN, DMB has collaborated with CCB to make the Data Interpretation System available. This facility provides the user with a transparent interface for retrieval of data from mainframe databases and enables reporting, display, spreadsheet, graphics and statistics on retrieved data at the LAN platform. To the user, all facilities appear to be the product of a friendly front end at the workstation.

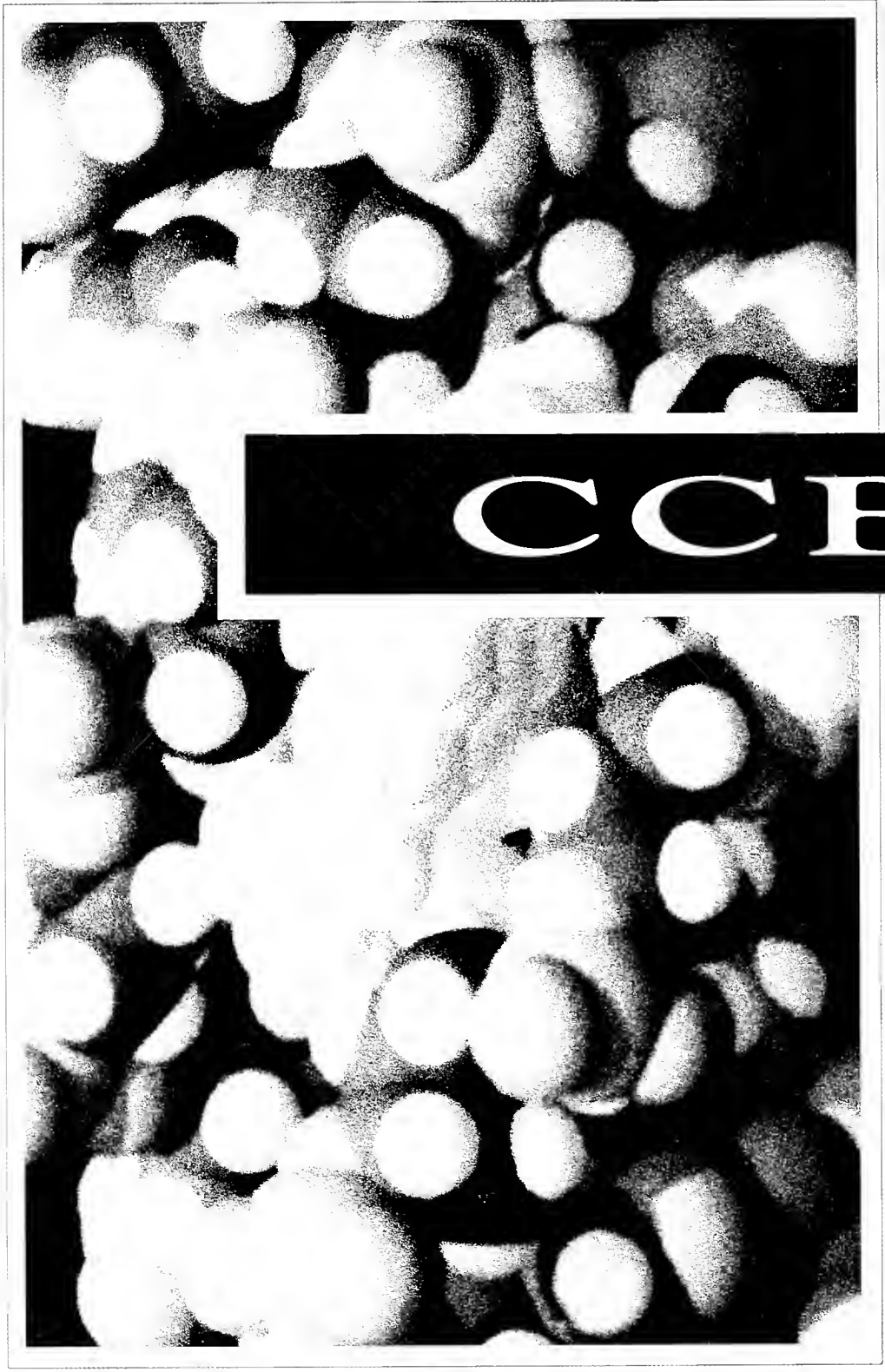
Future Plans

To complete the evaluation of personal computer RDBMSs, two applications have been selected for development using client server technology. DMB will move the maintenance functions of the Executive OD, Secretariat, correspondence control system to a test mode using the new technology; and a new system

in the Medical Records Department will also be implemented. If this proves successful, the Executive Secretariat will be fully converted and a process to use the client server approach to support all OD applications and distributed processes can begin.

During FY92, DMB plans to convert selected subsets of the appropriate ADB data to a relational database environment and to begin the development of improved tools for searching and analysis for the ICDs. This effort will require 20 to 24 months. During that period, we will also be evaluating advances in technology to determine whether and when full conversion might be recommended. Computer Assisted Systems Engineering (CASE) technology will also be used in the conversion process from IMS to DB2.

We believe that now is the time to consider all opportunities for distributed processing. However, we have concluded that it would be premature to consider "distributed database" as a viable alternative to shared central database for large-scale, mission-critical enterprise data such as the ADB. It is also clear that fully distributed database technology very possibly could develop along a faster path than a viable alternative to mainframe relational technology. With this in mind we feel that the best approach is to take advantage of existing, well-understood, validated database approaches for the next few years. In this way, we can make advanced capabilities available to the NIH community without encountering the risks of an unstable environment.



CCB

Computer Center Branch

William C. Mohler, M.D., Acting Chief

Executive Summary

The Computer Center Branch has engaged in a number of initiatives to improve its management and service to its customers. For example, CCB has begun a substantially improved capacity planning program for its mainframe IBM 370 systems.

Support for scientific computing expanded, with several new applications programs for performing computational chemistry, mathematical analysis, and graphics programming on the Convex system. Consulting and training services expanded during the year, reaching a larger audience with a broader range of technologies.

NUnet, the integrated network connecting campus LANs, the Central Computer Utility, and international networks, has been improved by the installation, configuration, and testing of system monitoring workstations, which are used continuously to detect problems before they affect service. In addition, a new effort was begun to establish a client-server database management system (DBMS) that will allow users of workstations or LANs to access the powerful database management capabilities of the mainframe.

System software improvements in FY91 contributed to increased user productivity by permitting high-speed printing of PostScript documents, connectivity for Macintosh computers, and easier PC-to-mainframe file transfer. In addition, the Computer Center's electronic bulletin board system, ENTER BBS, grew steadily, with two major enhancements added to improve conferencing capabilities.

Of historical note was the February 1991 phaseout of the DECsystem-10 timesharing computer, which had provided computational support to the NIH research community for over 22 years. With help from CCB staff, most DECsystem-10 users transferred their work to the Convex.

The Computer Center Branch designs, implements, operates and maintains the central NIH Computer Utility and its associated telecommunications and networking facilities. The Utility is made up of two interconnected multicomputer facilities designed around large-scale IBM and Convex mainframe computers. It has evolved over the past 2 decades to provide computational and networking services in support of the dynamic and diverse requirements of NIH research investigators and administrators who conduct and manage modern biomedical research.

The Utility provides interactive timesharing, wide area networking, database management, graphics, batch, high-performance scientific computation (e.g., vector processing), and word processing services to approximately 17,500 authorized users at NIH and in 29 agencies throughout the federal government. The Computer Center Branch provides its services on a fee-for-service, full cost-recovery basis.

Services of the Utility are available worldwide through the TYMNET, and Internet international data communications networks as well as through the Federal Telecommunications Service (FTS), and commercial switched telephone services. Thousands of users on local area networks (LANs) are connected to the Utility via dedicated, high-speed communications lines and a campuswide area network. Access to BITNET is also provided through the Computer Utility.

Strategic Planning and Peer Review

CCB is an active participant and contributor to the development of a DCRT Strategic Plan, an effort initiated in FY91. Because the CCB has historically been the largest supplier of computing services and support available from DCRT, CCB has been able to provide a unique service-oriented perspective to DCRT's

Chart 1.

NIH Computer Utility Hardware

IBM 370 HARDWARE

The IBM facility is an integrated multiprocessor complex, interconnected by shared disk storage. The processors are five model 3090-300J's with 128 megabytes of main memory, and 512 megabytes of expanded memory. The peripheral devices include:

ILK 3722 Ethernet interface for TCP/IP
3380K disk drives
3480 cartridge tape drives (18 track, 38,000 BPI)
3422 tape drives (6250/1600 BPI)
3420-5 tape drives (7 and 9 track)
3800-3 laser printing subsystems
4245 impact printers
6670 scientific printer
3827 cut sheet printers
8010 Datastream protocol converters
3172 channel to Ethernet interface
3705 communications controllers
3725 communications controllers
3745 communications controllers.

Peripherals are available to all processors, providing non-idle redundancy and minimal disruption of service in the event of any subsystem or component failure. Offline output devices that accept tapes from the IBM 370 include the SD-4561 microfilm recorders and the 925/1055 CalComp plotter.

Convex C240 HARDWARE

The Convex C240 consists of four processors with hardware support for vector processing and for scheduling parallel processes. Each CPU has shared access to 512 megabytes of memory and five input/output processors. Input-output devices include:

DKD-308/208 & 502 disk drives
MTD-102/002 tape drives (800/1600/6250 bpi).

The Convex system is generally restricted to NIH use.

resource infrastructure planning efforts, particularly in the areas of interoperability, scientific computing resources, and information access through databases.

Coincident with the DCRT Strategic Plan efforts, the Division has initiated a formalized peer review process of all of its laboratories, branches and major program activities. CCB and the Network Task Group collaborated on the preparation, presentation and conduct of the Networking Peer Review held in June 1991. The DCRT peer reviews are patterned after those established for the NIH Intramural program.

Capacity Planning

CCB has begun a major and substantive enhancement of its capacity planning effort for its IBM 370 mainframe systems. Over the last 5 years, there has been a movement of some types of processing away from the mainframe platforms to the LAN and PC platforms. This shift in workload contributed to the presence of excess capacity on the installed base of System 370 mainframes. To correct this situation, CCB removed one of the 3090-300J mainframes from service at the Utility.

In order to better measure and predict processing capacities against current and projected workloads, CCB will utilize capacity planning software (e.g., BEST/1 of BGS, Inc.) that employs sophisticated modeling algorithms that can be calibrated to specific hardware, software and operational environments. The use of capacity planning tools will complement an installed base of performance measurement and tuning tools (which include OMEGAMON, MICS, RESOLVE, etc.). In addition, we have systematically altered capacity in an experimental fashion and measured the effects on histograms of response time. Similarly, we are evaluating the capacity of other major system components. As a result of these analyses, we have cancelled two previously scheduled optional upgrades.

Capacity planning for the IBM System 370 complex is only part of the total Information Resource Management (IRM) initiatives being addressed by CCB. Evolving technologies such as distributed processing across heterogeneous platforms create challenges in capacity planning that reach well beyond the domain of mainframes.

NUnet Revolutionizes Scientific Communications at NIH

It has been just 18 months since the Computer Center Branch, together with the Network Task Group and other DCRT components, began its full-scale effort to provide electronic connectivity among all NIH LANs both on and off campus. The expanding NUnet has provided NIH scientists, administrators, and programmers with one of the most comprehensive and sophisticated LAN communications networks available anywhere. (See Chart 2 on next page).

NUnet is an integrated wide area network designed to interconnect LANs, the Computer Utility mainframe computer systems, and international communications networks. A high-speed network of dedicated fiber optic and copper conductors, NUnet is providing state-of-the-art connectivity for thousands of users throughout NIH, and has become the foundation for electronic communications among members of the NIH community, outside academic institutions, and research organizations around the world.

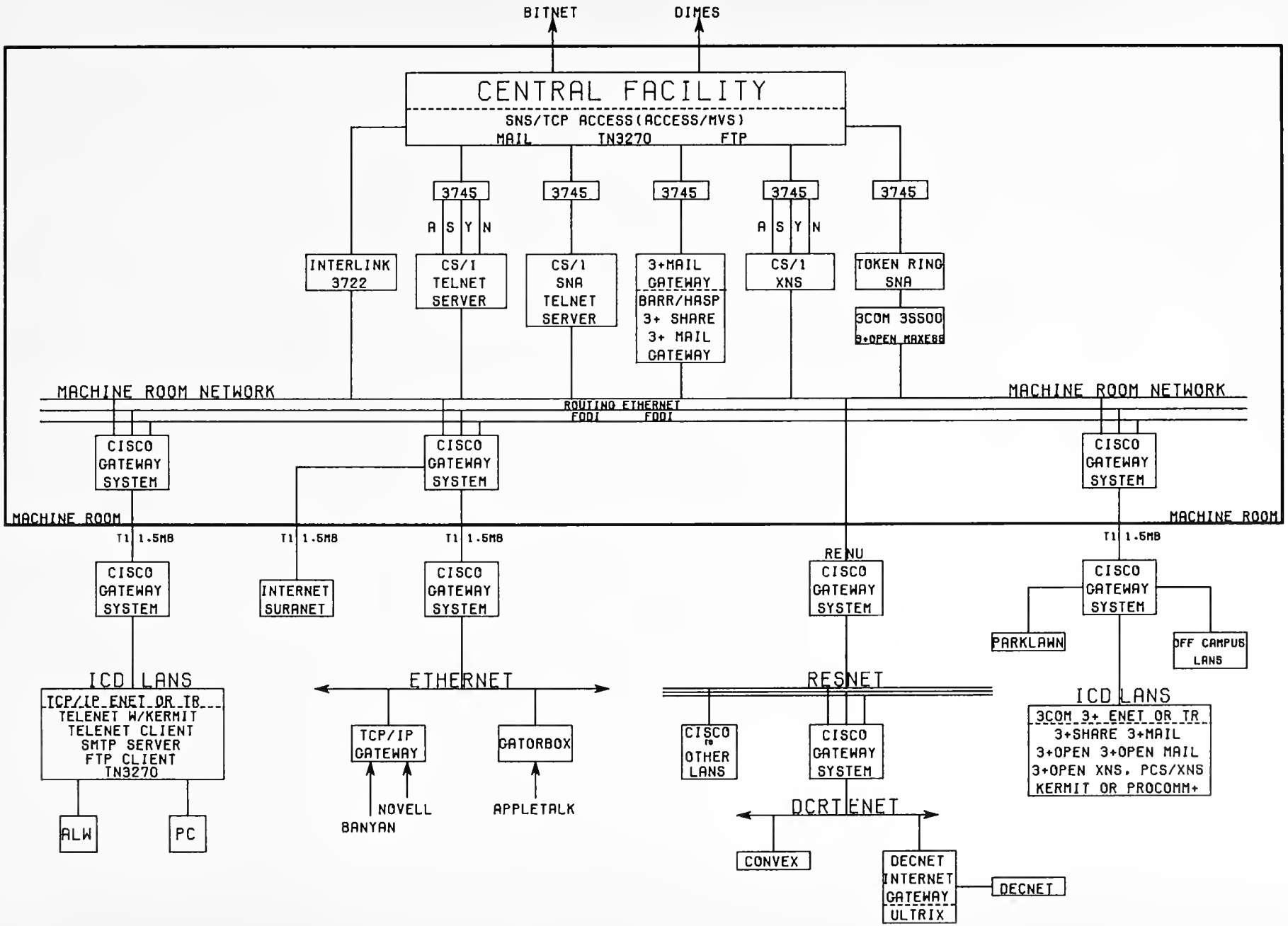


Chart 2. NIH Computer Utility Network Connectivity

Over 60 LANs have been directly interconnected and provided with high-speed access to the Computer Utility's System 370 and Convex mainframe computers and to other computers and workstations on the Internet and BITNET international data networks. Connected LANs are located in every ICD of the NIH, on and off the NIH campus, including NIEHS at Research Triangle Park in North Carolina. Access to the NCI Cray supercomputer in Frederick, Maryland and to the NIH RESnet (including approximately 30 LANs) is provided through NUnet.

NUnet uses a full-duplex, dedicated-circuit star topology radiating from the Computer Center's Routing Ethernet. Reliability, flexibility, and maintenance are hallmarks of the system, which offers communication speeds from 20 to 50 times faster than those previously available. The initial implementation offered communication speeds of 1.5 megabits per second. Ultimately, speeds of 100 megabits per second will be supported.

A network as expansive as NUnet requires extensive network monitoring facilities; thus, a major undertaking of the Computer Center networking staff during the past year was the installation, configuration, and testing of Proteon Overview network monitoring workstations. These are used continuously to detect NUnet problems before they affect service. Many router and telecommunications line problems have been isolated and fixed before the users of NUnet even knew a problem existed.

High-speed access via NUnet to the Administrative Data Base and other data retrieval and searching applications residing on the NIH Computer Utility became available during the past year. Another new facility, termed "anonymous FTP," allowed NIH users to make data files, research papers, and other information available quickly and easily to other researchers at NIH and thousands of other institutions in the world that are connected to the Internet. Anonymous FTP allows people at NIH or at other institutions on the Internet to access files stored on the CCB mainframe computers even if they are not registered users of the NIH Computer Utility.

Already, NUnet is changing the way scientific researchers communicate. For example, the NIH Guide to Grants and Contracts, which averages 18 pages in printed form and requires weeks to print and distribute through regular mail, is now being transmitted at electronic speed to more than 250 institutions throughout the United States. Researchers can receive the EGuide days or even weeks before the paper version is available.

Scientists engaged in intramural research have also enjoyed the advantages of the worldwide connectivity provided through NUnet. It is now common practice for NIH researchers to transmit drafts of collaborative research papers for comment and revision to colleagues at other institutions instantaneously. The draft-comment-rewrite cycle has been shortened from weeks to days or even hours.

With Internet, high-speed transfer of data and computer programs between workstations and computers on the network is now possible. The NUnet connection to the Internet allows NIH researchers to transmit or receive massive amounts of data in a very short time—something that was impossible just a few years ago. By using "anonymous FTP," both the intramural and extramural research communities can now collect and distribute research information and data worldwide at electronic speeds.

High-speed file transfer, remote job submission, and mainframe access via Telnet and 3270 connections are available via the versatile communication protocol TCP/IP. In order to access TCP/IP services on a NUnet-connected LAN workstation, a compatible communications software package must be

installed on the workstation itself. As a service to users of the NIH Computer Utility, the CCB has acquired a site license for PC/TCP, a product from FTP Software Inc. for personal computers that run DOS. The CCB offers Level 2 support for PC/TCP. With the site license, NUnet-connected users may obtain PC/TCP directly from the CCB.

This product opens the door for effective use of the high-speed connectivity capabilities provided by NUnet. The availability and utilization of this high-performance product is an example of the benefits that can be achieved through a collaborative effort between technical LAN coordinators and the CCB staff. Software acquisition and distribution (along with other networking services) has been a challenging addition to the responsibilities of the CCB staff.

ENTER BBS Grows

The Computer Center's ENTER BBS electronic bulletin board system, which became available at the beginning of FY91, proved enormously popular as a mechanism for sharing and disseminating information throughout NIH and the Nation. The number of bulletin boards grew steadily throughout the year. Currently there are 24 "public" boards that can be seen by anyone who enters the system, while 26 others are "hidden" because they are restricted for the use of a selected group or are still being developed. (During its first year, ENTER BBS was accessed more than 20,000 times.)

The adaptability of ENTER BBS to the various needs of diverse groups of users is a big reason for its success. Scientists, federal employees, managers, administrators, and computer specialists have all found ENTER BBS a useful source of information. Designed to be used by people who are not computer professionals, ENTER BBS avoids computer jargon and provides three levels of end-user interaction to accommodate the full range of user experience.

As anyone with access to the NIH Computer Utility can use the public bulletin boards contained within ENTER BBS, it has grown to contain information made available by people from many different organizations and professions. More than 20,000 times individuals at NIH, across the Nation, and around the world have used ENTER BBS to read online information, download files, and participate in electronic conferencing. New bulletin boards have been added regularly as users discovered the value of this exciting new facility.

As a result of the flexibility and popularity of ENTER BBS, the facility took over the functions of the old ENTER EXCHANGE in July 1991. To make EXCHANGE information available to the widest audience possible, all information contained in the ENTER EXCHANGE was moved into an EXCHANGE bulletin board of ENTER BBS, and the ENTER EXCHANGE system was removed from service.

Two major enhancements were added to ENTER BBS during the year. The first allowed users to see whether there are unread messages in any conference for which the user is registered. Previously, it was necessary to join all of the conferences to see which ones contained new messages. The new feature shows information about new conference messages for all public bulletin boards, not just the one in use when the command is issued. A second enhancement offered the ability to link a conference to one of the discussion groups that exist at many of the BITNET nodes around the world. These are collections of individuals (and their electronic mail addresses) who are interested in discussing a specific topic. As in electronic mail groups, discussions are usually public and deal with a specific topic. There are over 2,000 lists of

varied discussion topics available. These enhancements to ENTER BBS were suggested by users of the facility; additional ones will also be based on direct user input.

Software Changes Make System More Versatile

FileAID, a COMPUWARE Corporation software product for ISPF/PDF (Interactive System Productivity Facility/Program Development Facility), was made available in July with Level 2 support. FileAID, a collection of data and file manipulation tools with options for SPF (System Productivity Facility), extended edit, and batch, provided a number of programming enhancements for users of ISPF/PDF.

Menu-driven, full-screen, interactive, ISPF-like displays and split-screen mode function are among the attractive new features of FileAID. The program eliminates ISPF's record-length and file-size restrictions; it allows both online browsing and editing of VSAM and sequential data sets and allocation or deletion of VSAM files. FileAID not only reformats entire files, but it can also be used to extract and reformat subsets of files using selection criteria based on record counts and/or specific values in individual fields. Previously, the only way to do this was to write a program to perform the extraction. Because FileAID recognizes job control language (JCL) syntax and continuations, it can be used to apply global changes to JCL libraries either interactively or in the batch. With the availability of PC/TCP (Personal Computer/Transmission Control Protocol) to users on workstations connected to NUnet, many more users are able to take advantage of the enhanced features FileAID offers at T1 speeds of 1.5 megabits per second.

An additional service to the Utility printing offerings was introduced in 1991. PostScript documents can now be produced on the high-speed cut sheet printers at the NIH Computer Utility. One of the most popular formats for producing highly formatted scientific documents and graphic materials, PostScript has emerged as a *de facto* standard for printed output. It is actually a language for describing characters and graphic output in ASCII code.

Customers of the NIH Computer Utility use numerous word processing and desktop publishing packages that produce sophisticated output in PostScript format on IBM-compatible personal computers, Macintosh computers, and Unix workstations. Such documents could previously be printed only on attached PostScript-compatible printers, which tended to be slow and printed only on one side of the paper. Now, PostScript documents may be uploaded to the IBM 3090 computers via FTP, Kermit, or the 3Com 3+ mail gateway and can be printed on the NIH Computer Utility's high-speed model 3827 cut sheet printers, in single- or double-sided format on plain or 3-hole punched paper, at 90 pages per minute. The Computer Center is currently offering PostScript printing at Level 3 support. We expect to be able to upgrade support for this facility after further evaluation.

An all-catalogued environment is the new standard for the NIH Computer Utility's public and private direct access storage. Catalogued datasets relieve users of having to remember specific dataset locations for their files and improves system performance because the Computer Utility can shift file locations as needed for more equitable distribution of files without adverse effects. In the all-catalogued environment, users create new, catalogued disk files without selecting a particu-

lar disk. This technique, called "nonspecific allocation," has long been available to users of the TSO and WYLBUR interactive systems, but users who run batch jobs have been limited in this capability. As of May 1991, batch users can now select one of three categories of public disk storage (temporary, permanent, or low-cost managed storage system) and let the operating system select a suitable disk volume for the new file automatically. Catalogued files are subsequently accessed simply by specifying the file name and status.

A new facility for submitting batch jobs automatically was introduced in May at Level 1 support, in response to numerous user requests. With the new ENTER SUBMIT command in WYLBUR, users can automate repetitive tasks, such as producing monthly reports, backing up data sets or making regular tape copies, without repeated effort. One command schedules the job to be performed at a user-specified time each day, on a particular day of the week, or on a specific date. The user simply sets the job up and saves it as an online, catalogued data set, then signs on to WYLBUR, types ENTER SUBMIT, and answers a few questions. Thereafter, the job is scheduled for automatic submission at the specified time or dates, and continues until the specified ending date is reached or the user clears it from the batch submission system. ENTER SUBMIT is particularly valuable to users of the anonymous FTP, a facility for setting up a file directory into which colleagues on the Internet may upload files. With the ENTER SUBMIT command, users can automatically process files received via anonymous FTP.

A new version of the NIH adaptation of MSKermit was released during FY91. MSKermit 2.32a+NIH2, a software package that provides terminal emulation and file transfer capabilities between IBM PCs and compatibles and the NIH Computer Utility, was supplied free of charge to users. The new version, which currently receives Level 3 support, allows users to direct the output of WYLBUR commands to printers attached to their PCs. This ability, one of the most requested enhancements to WYLBUR, gives users the capability of producing letter-quality output right in their work environment rather than waiting for output from the Central Computer Utility to be delivered. Other improvements include fast page scrolling, improved screen clearing, and a larger scrolling buffer. The new release also offers improved support for accessing the Convex computer at the NIH Computer Utility. Scripts distributed with Kermit 2.32a+NIH2 allow users to specify an attached hard-copy printer. A new WYLBUR command, HARDCOPY, directs output to the printer attached to the PC. The HARDCOPY command supports special printing features, such as bold, underscore, italic, and half linefeed, for many attached printers. Users accessing WYLBUR with NIH8188s can also use the HARDCOPY command to direct output to printers attached to their terminals; and a HARDCOPY command has been added to ENTER MAIL to allow users to print mail on the printer attached to their PCs or NIH8188s.

Connectivity for Macintosh computers increased markedly this year as CCB implemented Level 2 support for VersaTerm, a communications package that gives Macintosh computer users dialup access to WYLBUR, TSO (Time Sharing Option) and the Convex system, and file transfer capability between the mainframes and Macintosh computers. This marked the first step in CCB's efforts to provide full connectivity support for the Macintosh. The Macintosh is widely used for both research and administrative activities at NIH. CCB is committed to meet the growing need for Macintosh connectivity services. The dialup and file

transfer capabilities now supported on the Macintosh are equivalent to those available to users of IBM PCs and compatibles via the Kermit and ProComm Plus software packages.

VersaTerm includes the Kermit error detecting and correcting file transfer protocol; thus, Macintosh users can upload and download files with 100 percent error-free transmission. Using VersaTerm, a user with a Macintosh and modem can connect to the NIH Computer Utility and use all the services that the Utility provides, including WYLBUR, TSO, Convex, the DB2 database management system, and IMS (for DELPRO). Macintosh users who connect to the NIH Computer Utility can also participate fully in the BITNET and Internet international electronic mail networks and in the ENTER BBS bulletin board system.

Easier PC-to-mainframe file transfer became available when the BARR/TRAN file transfer facility was installed early in FY91. BARR/TRAN, which transmits files between personal computers and the 3090 mainframe, is a feature of BARR/HASP, a software package on PCs that simulates a JES3 RJE workstation. BARR/HASP is widely used in CCB to transmit jobs to the Computer Utility and to receive printed output. The BARR/TRAN facility, which is used when transferring files to or from the mainframe and a PC running the BARR/HASP RJE package, provides automatic file format conversion between the most common file organizations used on both the mainframe and on PCs. Its installation facilitated the development and use of software applications that require the transfer of data from PC to mainframe and mainframe to PC.

Collaborative Effort Facilitates Client-Server Technology

A major new effort was begun early in FY91 that has as its ultimate goal the establishment of a client-server database management system (DBMS) at NIH. One of the most exciting and valuable computer technologies developing today, client-server technology combines the user-friendly capabilities of individual workstations (the client) with the powerful, production-oriented capabilities of mainframes (the server). By linking these two systems with high-performance connectivity, such as that provided by NUnet, an entirely new level of database service becomes possible.

A client-server database management system allows users utilizing individual workstations or LANs to access the powerful database management capabilities of the mainframe. Client software is selected from a variety of available programs to match specific user needs. The server handles database queries received from client products while performing the classic database management roles—handling access authority and data security, performing data backup and recovery, allowing data updates while guaranteeing complete data integrity, and ensuring high performance levels. Because all data are maintained by the server, reliability and stability approaching that of the operating system itself is essential.

DB2, the relational database management system which has been in production service since 1985, is a natural choice for client-server products running on the NIH Computer Utility. A robust and evolving product that has become the industry standard, DB2 is known for its reliability: in over 5 years of service at NIH, no data have been lost due to system failure! DB2 data are accessed via Structured Query Language (SQL), the industry standard interface for DBMS

access, thus ensuring compatibility with a wide variety of client products. Finally, CCB staff has developed the technical expertise needed for the success of a system-integral product like DB2.

Client-server technology has sufficiently great potential that CCB has begun to collaborate closely with colleagues in the Personal Computer Branch, the Data Management Branch, and the NIH user community to test several client products that may find use at NIH. A joint project is underway to evaluate various facilities for accessing DB2 from Macintosh computers and DOS-based PCs. Expertise was sought to select products likely to meet the diverse requirements of the NIH user community, and an invitation to join the evaluation process was extended to all the users of the NIH Computer Utility.

CCB recruited a group of technical lead database coordinators to work on evaluating PC and LAN products that connect to DB2 via NUnet. CCB staff worked closely with several vendors to set up presentations and software demonstrations in the NIH testing environment. Several promising products have been identified. The selected products, which are designed to enable users of PC and LAN databases to extract information from DB2 or allow simultaneous access of LAN and DB2 databases, are compatible with the operating system and NUnet communications environment in place on the NIH Computer Utility and seem to have compatible future goals.

The initial testing was designed to determine how the selected products work in the NUnet environment and to detect major deficiencies. Considerable headway has been made, and DCRT has established a major commitment to continue collaborating with the NIH user community in order to select multiple client products that meet the needs of all users of the mainframe relational database "engine."

CCB, together with PCB, DMB, and OD, DCRT, is actively cooperating with the Division of Research Grants (DRG) in its effort to migrate its current IMPAC (Information for Management Planning Analysis and Coordination) system to a multiplatform environment utilizing client-server technologies and relational database management systems. The IMPAC Migration Project has been active for over 2 years, with several years of work remaining to be done. CCB's role in the Migration Project is to be that of a consultant and facilitator. CCB has committed resources to the installation of hardware and software (including network connectivity facilities) required for a wide range of feasibility tests using products such as Oracle, Sybase, and DB2.

New Applications Support Scientific Computing

Several new applications for performing computational chemistry, mathematical analysis, and graphics programming on the NIH Convex system were introduced early in FY91.

Gaussian 90 performs semiempirical and *ab initio* molecular orbital calculations to model the structure and dynamic behavior of molecules.

ML/e (pronounced "Emily") is a powerful and efficient system for performing calculations, mathematical modeling, numerical analysis, and graphics. One of the successors to the DECsystem-10 MLAB package, ML/e features intuitive, ready-to-use commands and a programming language to customize commands for special applications, including differential equation solutions, curve-fitting, and parameter estimation for models.

FIGARO is a device-independent implementation of the Programmer's Hierarchical Interactive Graphics System (PHIGS) standard. Designed for 2-D and 3-D interactive graphics applications requiring

hierarchical data structures, geometric modeling, and interactive input, it allows users to create their own PHIGS applications using FORTRAN or C programming languages.

USENET News is a worldwide bulletin board on which users of Unix systems exchange ideas and information on molecular biology, computer technology, and a wide range of other scientific issues.

CHARMm models the structure, dynamic behavior, and characteristics of biologically relevant molecules using empirical energy functions, incorporating the effects of solvent, temperature, and experimental data into its simulations.

The installation of the latest release of the CHARMm molecular modeling package on the NIH Convex system provided users with a number of significant enhancements. An important new feature is the ability to offload portions of computation-intensive problems from local workstations to the Convex computer, thus enabling researchers to distribute problems to the platform where they can be solved most efficiently.

QUANTA, a molecular graphics analysis program that runs on Silicon Graphics workstations, displays full-color, three-dimensional images of molecules that can be manipulated interactively and visualized in different ways. QUANTA also provides user interface to CHARMm and several other programs. The network version of CHARMm from Polygen Corporation enables CHARMm calculations to be executed on other computers while QUANTA continues to run on the workstation. CHARMm calculations are performed on the local workstation by default, but for improved performance when doing computation-intensive calculations such as molecular energy minimization or molecular dynamics calculations, CHARMm can be executed on the NIH Convex system.

Distributed computing at NIH has a bright future. Research in this area has been going on for years, resulting in the introduction of products like QUANTA and CHARMm with increasing frequency. The recent surge of interest in scientific visualization techniques has spawned several promising projects that capitalize on the modular structure of programs and network communications between computers. The Convex staff is examining biomedical applications to see how the NIH research community might benefit from this new distributed processing technology.

DECsystem-10 Retired

The DECsystem-10 timesharing computer was phased out of service in February 1991, having provided computational support to the NIH research community for more than 22 years. The end of DECsystem-10 timesharing at NIH was recognized with an open house for the NIH user community. The event featured tours of the machine room and exhibits on the many accomplishments of the DECsystem-10 and its staff.

Most DECsystem-10 users transferred their work to the new NIH Convex supercomputer during the early months of FY91. While users were responsible for transferring data and applications to other systems, CCB staff were instrumental in making the transition a graceful one. Several key applications were installed on the Convex system and detailed suggestions for transferring DECsystem-10 applications to other computer systems were made available.

Because the MLAB modeling laboratory proved to be indispensible for researchers at NIH, two new programs with the functions of the original MLAB were made available. ML/e was installed on the Convex system to offer mathematical modeling in a timesharing environment, and PC-MLAB was offered for personal computers, with support from LAS and LSM.

Computer Assistance and Documentation

For over 23 years, the Programmers Trouble Report (PTR) system has served as a means for users of the NIH Computer Utility to communicate with the Computer Center to report technical problems, request refunds, suggest enhancements to existing services, and request new facilities on the IBM system 370 mainframe. This vital communication system received a major upgrade during FY91 to make it even more efficient and useful. Even the name of the system was changed—what used to be know as the Programmer Trouble Report system is now the Problem Tracking and Reporting system. PTRs have now entered the electronic mail system to offer convenient access to even greater numbers of people. As of October 1990, electronic mail could be used to report a problem, make a suggestion, or ask a question. Responses to PTRs from CCB staff were returned via electronic mail. Newly submitted PTRs recorded the user's electronic mailing address, allowing users to receive convenient, timely responses at any address on a NUnet LAN. Electronic mail responses were received days sooner than they had previously.

Electronic mail allows PTRs to be submitted more easily than ever before, opening up direct lines of communication to any user on a NUnet LAN, regardless of whether they are registered for WYLBUR services. LAN users were particularly encouraged to suggest improvements in NUnet service and report problems with NUnet, so that services and facilities could be enhanced to better meet their unique needs.

Input from users has always been vital to the function of the Computer Center Branch, and the enhancements to the PTR system are an important step in CCB's continuing efforts to increase communication with the NIH scientific and administrative computing communities.

This year, a total of 2,458 PTRs were received and answered. Extended assistance was provided to 195 individuals and groups through consulting appointments. More than 23,000 regular requests for customer assistance were received. Changes to improve the performance and reliability of the operating system required the implementation of 18 software reconfigurations. Approximately 26,000 fixes, both preventive and corrective, were tested and applied to the system and 16 new releases of current software packages were installed.

Documentation for the Computer Utility underwent dramatic change this year to enhance user convenience. The weighty Computer Center Users Guide, which describes all standards, facilities, and services for the IBM System 370 mainframe facility, was reduced by 200 pages at its spring revision, and each of the four language sections—FORTRAN OS/VS COBOL, PL1, and Assembler Language—were moved to individual manuals. The new programming language publications were made part of the manual sets for each language and sent automatically to users who subscribe to those sets.

The Users Guide was also reorganized, and the expanding networks section was moved forward in Volume 1 in order to reflect the growing importance of network connections in contemporary computing and the growing number of NUnet services.

There were 29 new technical documents published and 19 revisions or updates. The Automated Documentation Service was used 58,196 times and 48,393 copies of technical publications were sent to those users and to individual requesters. INTERFACE, the Computer Center's technical newsletter, published five editions plus the annual index.

Training Program Offers Varied Opportunities

The DCRT Computer Training Program offered many popular courses and seminars from previous terms and numerous new offerings that reflected recent advancements in high-speed networking. Altogether, 96 different classes were given by the DCRT Computer Training Program during the year. More than 3,200 students were accepted for classes. Fifty-eight short seminars helped to meet the diverse need for training among users of the NIH Computer Utility. Popular seminar topics included the NIH centralized bulletin board system (ENTER BBS), NUnet, mainframe mail connectivity, using the Internet and BITNET networks, the Unix operating system for scientific workstations, and seven Convex-related topics. An overview of the procedures and services of the NIH Computer Utility was offered for the benefit of newcomers. Other seminars dealt with software products, such as the Emacs screen-oriented text editor, DNAdraw software for the PC, and Macintosh shareware, and scientific topics, such as quantum mechanics, image processing, sequence analysis, laboratory data display, computational folding of proteins, and cluster analysis.

Selected Courses and Seminars in the DCRT Computer Training Program Coordinated by CCB

Seminars
Enter BBS
Network Services
Introduction to CASE
Advances in Tissue Optics
Flow Cytometry Topics
Laboratory Analysis Package (LAP)

PC
Windows Sampler
Cracking DOS 5.0's Shell
PC-DOS Advanced Topics
Memory Management on the PC
Running SAS Software in the PC-DOS Environment

Macintosh
Macintosh Software for the Scientist
Using MacVector
Macintosh Networking with TCP/IP

IBM 370
Introduction to ISPF/PDF
Beyond Basic WYLBUR
Introduction to the SPSS at NIH
Designing Tables and Managing a DB2 Database
Use of SAS at NIH

Unix
Fundamentals of Unix
Convex Questions and Answers
GCG Sequence Analysis on the Convex
Andrew File System

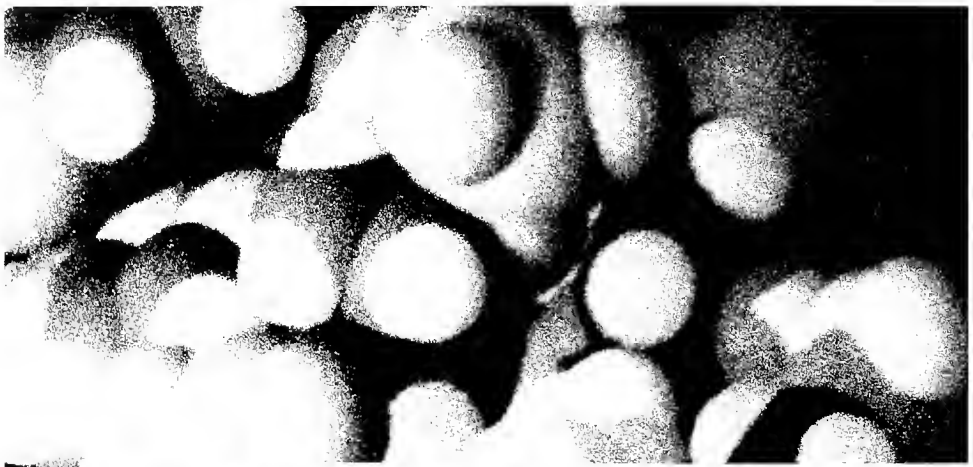
Self-study courses provide an alternative or supplement to the classroom courses offered in the DCRT Computer Training Program. A total of 44 different self-study courses were made available encompassing a wide variety of computer topics. The new Crwth computer-based training system added last year proved highly successful. A full-screen self-study system, Crwth (rhymes with truth) enables users to take courses at their own pace directly from a terminal or personal computer under TSO. The following Crwth courses were first made available in FY91:

- Using SQL/QMF: Basic Reporting
- Advanced SQL Using QMF
- TSO REXX Application Programming
- Using FOCUS: Basic Reporting
- Programming ISPF Dialogs

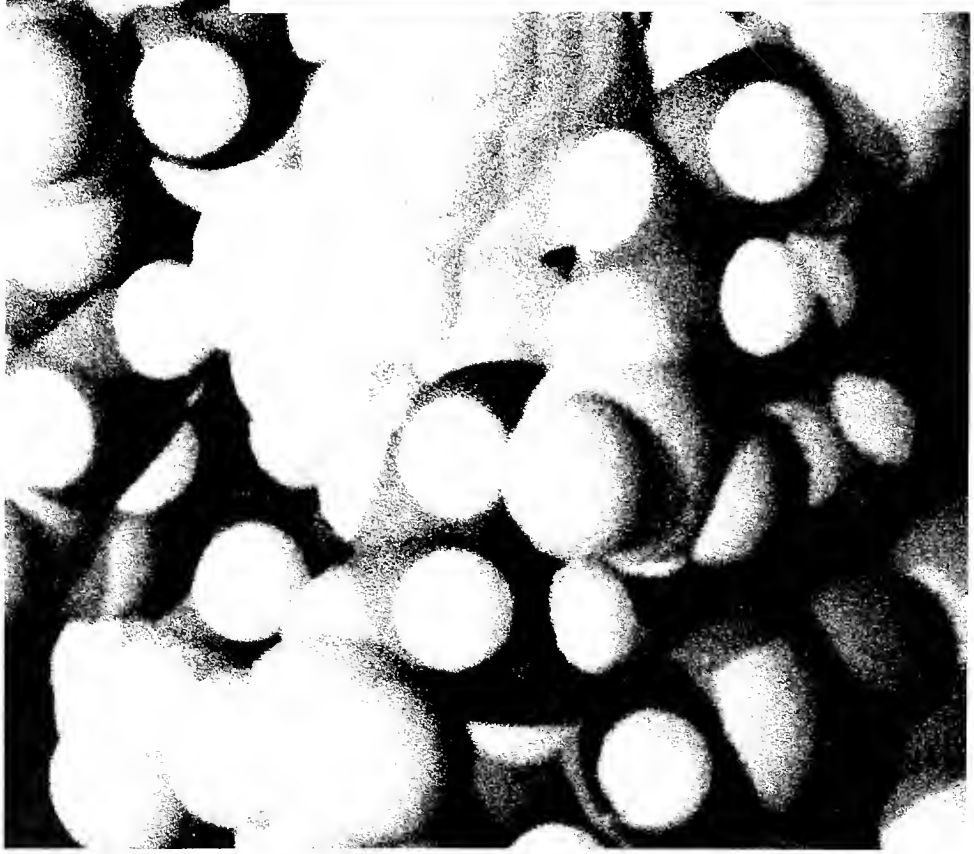
The DCRT Computer Training Program now boasts a total of nine Crwth courses taken by 156 students. A total of 26 traditional self-study courses were available and more than 250 students took advantage of this option in FY91.

Traditional self-study courses were also important. Self-study courses typically utilize texts, workbooks, and practice exercises which may be done on a computer. Two new self-study courses in the C language were added in FY91: "Coding in C Language" and "Advanced C Language Coding." These courses supplement the introductory self-study course in C Language.

The eight ABC (Assisted By Computer) courses remained popular. Interactive ABC sessions held during the year numbered 9,600, and ABC courses were completed by 876 people.



OD



Office of the Director

David R. Rodbard, M.D., Director

The Office of the Director (OD) provides overall program and management direction for DCRT. The Director, Associate Director, and Executive Officer work together as the immediate Office of the Director, whose activities in FY92 encompassed such issues as:

- strategic planning
- the High-Performance Computing and Communication initiative
- information resources management
- capacity planning and management
- peer review
- liaison with other Federal agencies
- management of the Division's computational biosciences program.

Two other offices supplement the work of the DCRT laboratories and branches:

The Office of Administrative Management (OAM) provides administrative and managerial support for the work of DCRT. OAM includes the Administrative, Personnel, Financial Management and Project Control Offices.

The Office of Scientific and Technical Communication (OSTC) serves as the central source of information about DCRT activities and about computing and related disciplines. It includes the DCRT Library and the DCRT Information Office.

Also within the Office of the Director is the Equal Employment Opportunity Office (EEO), which manages a full EEO program for the Division. The office serves as the focal point and advisory for all activities relating to the equal employment opportunities of DCRT employees and applicants. The EEO Officer maintains a close working relationship with the NIH Division of Equal Opportunity and other components concerned with minority and women's issues.

The division's role in NIH Information Resources Management (IRM) activities is shared among members of the immediate Office of the Director, but responsibility for NIH ADP planning, budgeting, and security resides in the Policy Coordination Section.

The Office of the Director also has a computational biosciences research component (described in the Research, Development and Service Section) that applies computing technology to biomedical research to solve problems in macromolecular structure representation and modeling for collaborating NIH scientists.

Equal Employment Opportunity Office

Gloria Richardson, EEO Officer

This year the DCRT Equal Employment Opportunity Office presented a formal introduction of the EEO Program to staff at the Division's first annual awards ceremony. Two supervisors and one staff person received EEO recognition with the award of plaques for their accomplishments, individually, in the support and promotion of equal employment. Two EEO counselors were selected to represent DCRT in NIH's attempt to resolve EEO complaints at the informal stage; a representative to the NIH Asian/Pacific Islander American Advisory Committee was selected; and the Division has established a fully chartered and active EEO/Employee Committee.

The DCRT EEO Office participated in minority job fairs and has been exploring graduate and alumni database systems for future recruitment resources.

The office also plans to conduct:

- seminars to prepare all DCRT employees for better equal employment opportunities
- EEO training for managers and supervisors
- training for prevention of sexual harrassment
- cultural diversity awareness sessions
- meetings with the private sector for outreach recruitment.

Acting through the EEO Office is an important channel of liaison and advice between the Director and all Division employees: the DCRT EEO Advisory Committee. Chaired by Margaret A. Douglas, the Committee encourages equal opportunity without regard to race, color, disabling condition, religion, sex, age, or national origin in all employment practices, and assists the EEO Officer in monitoring these practices. The Committee informs the DCRT community of EEO policies and how these policies are being implemented through the DCRT Affirmative Action Plan.

In FY91, the Committee was active on a variety of fronts. As a result of a Committee initiative, parking spaces for employees with disabilities were moved closer to the Building 12 complex and increased in number. A shuttle bus stop was also relocated to bring employees closer to the building.

The Committee sponsored several lunchtime seminars to help DCRT employees improve their skills, and an employee profile board was renovated and regularly updated with pictures of various DCRT groups to increase internal communications.

In the area of outreach, Committee members began negotiations with the NIH Division of Equal Opportunity to cosponsor a promotional diskette holder featuring a message about hiring persons with disabilities. A Committee effort is also under way to participate in the NIH Adopt-a-School program, which would give DCRT employees the opportunity to mentor students at a local high school.

Office of Administrative Management

Marian Dawson, Chief

The Office of Administrative Management (OAM) provides guidance and support on all administrative and business management aspects of the Division's programs, advising on the management of resources, the provision of administrative services, program planning and evaluation, and policy and legislative analysis. The office is organized functionally in sections for administration, finance, and personnel.

During this past year, the Office of the Chief has been actively involved in the DCRT Strategic Planning Initiative, organizational planning, review of administrative services and processes, and implementation of the Procurement Integrity Act. An effort was launched to review administrative and budget reporting in order to streamline reports being produced.

The *Administrative Management Section* advises and assists management in the full range of administrative services, including procurement, property, travel, space, and other support services.

During FY91 the Administrative Management Section has been involved in the implementation of several NIH online systems. DCRT was part of the NIH-wide property inventory, conducted by onsite contractors. This system went online in the spring of 1991 with the requirement that all property transactions, i.e., acquisitions, transfers, and property surplusing, be entered into the system. Another online system introduced this year was the preparation of travel vouchers.

During FY91 the office initiated quarterly secretarial meetings to convey to staff the many changing requirements for the processing of administrative services.

In response to audits of timekeeping practices at NIH, the Administrative Management Section conducted a series of reviews with DCRT timekeepers of the practices and procedures of the timekeeping function in the Division. These reviews were conducted to identify timekeeping weaknesses and to bring DCRT

into compliance with timekeeping requirements.

The *Budget Office* continued to carry out its annual budget functions for DCRT. In addition, materials were prepared to orient the new Division Director on financial matters. The Office also investigated and assessed alternative methods of funding for specific servicing activities and provided information for the Ad Hoc Peer Reviews.

Staff in the *Project Control Office* opened 194 new accounts and registered over 1,700 new users during the year. In addition, the office participated in setting up policies and procedures to facilitate usage of the Convex facility and registered over 500 people to use that facility. The office also completed its annual update of information on over 3,400 accounts and over 20,500 users currently in the Project Accounting System.

The DCRT *Personnel Operations Section* advises and assists management in providing and optimally utilizing human resources to accomplish divisional goals; it is also responsible for conducting the Personnel Management Program for the Division. This includes staffing and recruitment services, employee benefits programs, retirement, training, employee relations, conduct and ethics.

The office initiated FY91 with the arrival of the new Director and responding to his new initiatives that included hiring at very active pace aided by an increasing number of special hiring authorities, especially for intramural scientists and college students. Personnel office staff attended OPM, NIH, and commercial job fairs and contacted a number of local colleges and universities in an effort to attract entry-level professionals. The office has been conducting a detailed analysis of organizational structure, developing appropriate recommendations. Plans for SES and other high-level recruitment are being developed.

DCRT was able to augment staff by utilizing the summer program. Approximately 25 university and high school students were hired for technical, professional and clerical positions throughout the Division. Creative staffing efforts such as the Stay-in-School Program (in which DCRT's participation doubled), the Federal Junior Fellowship Program, and the cooperative education program have helped increase the number of minority employees within the Division.

The Personnel Officer instituted a seven-point Total Human Resources Program in an effort to address a wide range of personnel issues in an organized approach. This program has already resulted in a greatly expanded awards program, which included the first DCRT Awards Ceremony; a greater emphasis on handicap, minority and student recruitment; and a greater emphasis on supervisory responsibility and development. The program, when fully in place, will become a continuing focus of the office.

The staff continued to provide an increasing variety of assistance and information on employee benefits and services including hiring mechanisms for entry-level professionals, new classification series and guides, and several additional award programs through seminars, meetings, status reports, memos, and follow-up interviews. The office also used several additional software packages, including one developed by DMB, to further automate personnel processes. Staff attended specialized training courses, seminars, and professional conferences to sharpen skills and keep abreast of new developments in human resources. Conferences attended included the NIH Human Resources Management Conference in November 1990 and the International Personnel Management Association's Montgomery County Chapter Spring Conference in April 1991 and Eastern Regional Conference in Wheeling, West Virginia in June 1991.

DCRT Library

Ellen Moy Chu, Head

The DCRT Library provides information resources in computer science, mathematics, and statistics, along with computer applications in biomedical sciences, engineering, information science, and management. As an active participant in the work of the Division, it pursues collaborative testing of new commercial computer applications for libraries, integrates useful offerings into the local area network (LAN), and uses a full range of telecommunication-based facilities and internetworking to communicate with its users and other libraries.

The multimedia collection and information services support other parts of NIH as well as DCRT. This year, over 61 percent of new borrowers were other NIH staff. Library staff processed acquisition or interlibrary borrowing requests of 4,775 books, reports, articles, journal issues, and subscriptions.

Library resources continue the migration trend from print to digital formats for information distribution. This year we added six subscriptions and two reference monographs in CD ROM or diskette formats. Four of these include network licensing and will be available to NIH LAN users. Licensing was renegotiated to increase access to two CD ROM publications, and to convert a CD ROM abstracting and indexing service from stand-alone to network licensing. These information systems, designed for end users, continue to decrease the number of staff-mediated online searches. Such searches comprise accesses to remote databases unavailable to end users and sessions requiring specialized training or expertise.

With inauguration of PUBnet, the Campus Users Research Exchange Public Network, a major change in LAN access to library resources began. Library staff worked with the PUBnet administrator to implement campus-wide access to library information. Selection of the DCRT Library menu choice provides access to

- the online catalog
- contents of the library brochure
- information about journal holdings
- listings of books purchased on approval available for users' review at the library
- the most recent issue of the monthly bulletin
- CD ROM publications.

NIH staff access PUBnet via NUnet, RESnet, or dial-up connections, and are thus able to conveniently consult library information systems from their offices and laboratories. Desktop computers and software applications for the new digital media provide many new access points to retrieve, manipulate, and output data. Initially, the library distributed network drivers to users in accordance with licensing for the software networking solution for CD ROMs. Upon library recommendation of another CD ROM networking system, the PUBnet administrator acquired and installed this system to improve performance and to simplify operations.

During National Library Week, over 50 users visited the library for an open house on International Special Librarians' Day. Seven computers were set up for hands-on demonstrations of four CD ROM information systems, the online catalog, Grateful MED, and PUBnet. A new library brochure updating information about services as well as electronic mail and network features was printed in time for this event. Later in the year, the public workstations were upgraded. In addition to a new PC, a Macintosh was ordered for information applications using this architecture.

DCRT is working with the NIH Editorial Operations Branch to facilitate production of the Scientific Directory/Annual Bibliography (SD/AB).

The Chief Librarian participated in Division analysis of the process and drafting of recommendations to expedite data entry. Recommendations included proposal of a pilot project to make SD/AB information available online and for desktop publishing. The library acquired PC and MAC versions of commercial bibliographic management packages to demonstrate to NIH components for use in assembling SD/AB data.

As an emerging technology, CD ROM lacks standardized user interfaces and techniques. The Head Librarian participated as a voting member of CD-ROM Consistent Interface Committee (CD-CINC), a joint industry-government working committee of the federal Special Interest Group on CD-ROM Application Technology (SIGCAT). The final CD-CINC report, which identifies and defines 13 basic end-user functions, will be distributed industry-wide and to standards committees to promote standardization efforts of this man-machine interface.

The Head Librarian completed a year as Chairperson of the Special Libraries Association Scholarship Committee. In November, she led a panel discussion on Microcomputer System Selection at the Washington, D.C. Chapter Update '90 Workshop.

Future plans will focus on enhancing campuswide information services via PUBnet. A concurrent activity will be developing training sessions and aids to optimize use of these systems. Investigation of multimedia CD ROMs and development of access for non-DOS workstations are also planned.

Research Projects

Information Retrieval Systems on Local Area Networks

Ellen Moy Chu

The project is developing LAN access to DCRT Library information systems. With implementation of campuswide access via PUBnet, network licensing agreements were renegotiated to extend coverage and additional publications were acquired to broaden services. Consultation services were provided to many NIH staff selecting and installing CD ROM hardware and software. Future plans include investigation of multimedia CD ROMs, Macintosh and Unix workstation access to library network information systems, and further enhancements on PUBnet.

Publications and Presentations

Chu EM. Internetworking a Library Within an Enterprise Network. In Breeding, M., ed. Local Area Networks in Libraries: Case Studies. Westport, CT: Meckler (in press).

Chu EM. Microcomputer System Selection, panel discussion moderator and speaker. Update '90 Workshop, Special Libraries Association Washington, D.C. Chapter, November 14, 1990.

Policy Coordination Section

John M. Campbell, Head

The Policy Coordination Section has responsibility for NIH ADP policy coordination and ADP reporting functions. These include ADP procurement reviews; the Information Technology Systems (ITS) budget; ADP tactical planning, administrative systems reviews, inventories and hardware; Federal Information Processing Standards (FIPS); and ADP computer systems/facilities security. The section also provides the staff support to the Director, DCRT; the Associate Director for Administration, NIH; and the

Deputy Director, NIH, for Information Resources Management (IRM) activities. The reporting functions are coordinated with the ICD's Planning Coordinators, Systems Security Coordinators, Systems Managers and Administrative and Executive staff. The staff works closely with the NIH ADP Clearance Officer in the coordination of data reporting.

The systems security function includes an annual Computer Security Training and Awareness Program consisting of specific subject matter discussions and presentations featuring the latest computer security protection devices, virus protection modules, and the more important aspects of computer security. The Policy Coordination Section conducts Computer Security Awareness Days featuring lectures on potential causes of widespread personal computer viruses impacting the NIH community. An important film production depicting computer security at NIH is the basic information documentary on internal security controls and the role of the managers. Other DHHS and Federal components have used the film. A calendar depicting computer tips and individual advisory posters have been displayed throughout NIH as another means of making the employee aware of the need to protect the information that they are responsible for and/or have access to during the course of the workday.

A major objective is to comply with the regulatory statutes for systems reviews and risk assessments/analyses. The NIH Standard Risk Protocol (SRP) was developed to enhance the process for completing a risk assessment, risk analysis and system review required by the Computer Security Act of 1987. This personal computer application can also be used by management in the IRM review process. This automated process is designed to measure the level of security in an application system/facility and to provide management with guidelines for implementing cost-effective safeguards to improve the security risk level. The Standard Risk Protocol consists of several modules. The completed modules are currently in use at NIH, other DHHS components, and a number of other Federal agencies. The protocol has been accepted as a cost-effective approach to managing ADP security, and now includes a summary risk assessment module that collects and stores data relative to risk analysis, contingency planning, reviews, and certifications. Cross-summarization of the data relative to systems, facilities, and equipment provides management with the capability of reviewing organizations and correcting the vulnerabilities from a total management perspective. The protocol is now available on a 3.5-inch diskette.

A module used to determine a computer system's probability of infection by a computer virus will be completed in 1991. This feature will provide further protection to the systems manager with respect to sensitive data downloading from the mainframe or independently developed microsystems. A major development that is under way in the section is the Contingency Planning and Disaster Recovery Module for NIH facilities and application systems that will provide the systems managers with the insurance protection needed to survive a medium- to large-range disaster to the operations of the facility with serious, long-term interruption of service. This module should be completed in 1991. A system review module is designed to determine the effectiveness of a system's development and operation, and to determine the effectiveness of the output reports generated by the system. The module examines the level of compliance with GAO principles and standards on the effectiveness, reliability, and integrity of financial applications.

Information collection processes for the annual

Information Technology Systems budget have been automated. This allows for budget information from the ICDs to be collected, compiled, and updated more efficiently. NIH budget information is transmitted to the Public Health Service over the network. Transmittal of information in this manner lessens the chance of error and moves the information quickly between components. Inventory of systems and equipment are automated and updated on files either on the mainframe or on a personal computer. A major effort to update the information is under way. Information gathered from the risk analysis process supports the reporting requirements of other IRM functions.

DCRT Information Office

Raymond Fleming, Information Officer

Transition was the by-word for the DCRT Information Office of FY91. The November 9 arrival of a new Division Director was followed by the December arrival of a new Information Officer, ending an extended period during which office functions had been overseen by the Executive Officer.

Planning was undertaken to 1) effectively communicate the partnership of computers and biomedical research to the NIH administrative, intramural, and extramural communities, and 2) build and maintain the internal identity of the Division. Areas targeted for special emphasis included: corporate identity and image development, publications, writing and editing, media liaison, and special events. Adding substance to the vision were plans for a new corporate logo, color, and makeover; a four-color flagship brochure; a "family" of DCRT publications; and feature articles in NIH and trade publications on DCRT collaborations in biomedical computing.

Editorial and writing projects accounted for much of the office's activities. The 1990 Annual Report was completed and printed, and a new draft edited for the 1991 version. The office extensively revised Computing Resources, helped produce a new library information sheet, shepherded through a new NIH Directory of Image Processing Facilities, and updated the Division's submission to the Scientific Directory and Annual Bibliography, the International Report, the Postdoctoral Research Fellowship Opportunity Catalog and the NIH Almanac. Office personnel also increased their editorial services, helping out with documents involving peer review, strategic planning, and award justification.

The office made a concerted effort to better inform the NIH community of its resources and its commitment to partnership with the bench scientist. The NIH Record published front-page DCRT-penned feature stories on networking, networking applications, highly parallel computing, electronic grant application development, and a collaboration with gene-sequencing scientists. Most issues of the Record, in fact, contained at least one DCRT story or photo/caption related to DCRT special events, training, or personnel activities.

Building DCRT corporate identity became a major part of the office's 1991 activities. Information office staff members organized the highly successful Division picnic, the holiday party, the first Division-wide annual awards ceremony, a retirement celebration for Associate Director Dr. William Mohler, and the dedication of a conference room in honor of Dr. Arnold Pratt, who retired after 25 years as DCRT Director. Special wall displays were developed for the EEO committee and for letters of commendation from the NIH community, and photo montages of several DCRT events were mounted and displayed. A new recruiting exhibit was developed, and designs obtained for a new logo for the Division. The Office

also initiated the idea of a special disk holder to promote hiring individuals with disabilities, and worked with the DCRT EEO committee and the NIH EEO Office to bring this project to fruition.

Several office-assisted media stories appeared in the trade publications Federal Computer Week and LAN Magazine. The office also distributed a press release announcing Dr. Rodbard's appointment.

Finally, the office continued its tradition of service to its external and internal public through its many daily information activities. Staff members ably handled information, publication, and photo requests ranging from several to a score per week, tracked Division-submitted scientific papers, and regularly issued DCRT News Clips. To keep pace with computer-based advances in the public information arena, the staff completed courses in several desktop publishing systems.

U.S.G.P.O.*1992-342-299:60033

Discriminatio pplic-
able public lav no
person in the s of
race, color, na ex, or
age, be exclud d the
benefits of, or der
any program ı of
Computer Re erated
in compliance

Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland 20892